----------Work in progress. Forthcoming with Georgetown Journal of Legal Ethics----------

Existential Advocacy

John Bliss[*]

ABSTRACT

Lawyers have played a central role in a wide range of movements for legal and social change. This article investigates how the tradition of movement lawyering is being applied to what could almost be described as humanity's ultimate social movement—the movement to mitigate "existential risk," which is defined as risk of human extinction or other permanent destruction of future human value. Over the past two decades, an Oxford-based academic community of philosophers and scientists has been cataloguing existential threats, including threats arising from engineered pandemics, transformative artificial intelligence, runway climate change, nuclear war, and extreme natural disasters. Although such events may be relatively unlikely in the near term, this list of threats is expected to grow with technological advances. Awareness of this issue has sparked the founding of dozens of non-profit organizations that work to preserve the future of humanity. In just the past couple years, this movement has entered the realm of law and politics. They have already made extraordinary progress with the introduction of new legislation, court decisions, and international policy initiatives. These advocates have ambitious plans for their next steps. At the same time, the movement has faced vociferous criticism in the media for being influenced by wealthy donors and shifting attention away from current social injustices.

This article presents the first empirical study of the legal activism in this movement. It asks how these "existential advocates" approach the key questions faced by all social-change lawyering campaigns: (1) efficacy, e.g. how to have the greatest impact when addressing such a large-scale, uncertain, and abstract issue; and (2) accountability, e.g. how to faithfully represent the multitudes of potential future generations who are silent stakeholders in decisions we make today. Drawing on a two-year qualitative study embedded with this community of legal advocates, the article describes the development of a distinct model of social-change lawyering—the "priorities methodology." This model aims to maximize impact using formal processes for optimizing goals and strategies while minimizing cognitive biases. While this

[*] Assistant Professor of Law, University of Denver Sturm College of Law. I wish to thank my research assistant Juliana Todeschi. Invaluable feedback was provided by Catherine Albiston, Alan Chen, Scott Cummings, Meghan Dawe, Benjamin Eidelson, Noah Feldman, Bryon Fong, Calvin Morrill, and David Wilkins. I received helpful comments at meetings of the UC Berkeley Center for the Study of Law and Society, the Harvard Law School Center on the Legal Profession, the University of Denver Summer Scholarship Series, the Multidisciplinary Forum on Longtermism and the Law at Universität Hamburg, and the Law and Society Association. I am grateful to Christoph Winter and other members of the Legal Priorities Project who granted me access to conduct an in-depth ethnography of their organization. I also wish to thank the 53 participants who voluntarily participated in research interviews. Funding for this study was provided by the Hughes-Ruud Research Professorship at the University of Denver. The study was approved by the University of Denver Institutional Review Board.

model is supported by some recent recommendations from socio-legal literature, it faces limitations when seeking to persuade legal decision-makers and incorporate a more diverse range of voices. The article concludes with recommendations for adapting this model as the existential risk community scales up and pursues more direct and high-profile legal interventions.

TABLE OF CONTENTS

INTRODUCTION 4
I. Background: 12
A. Existential Risk 12
B. Literature on Law and Social Change 17
II. Research Design 19
III. Theory of Efficacy: The Priorities Methodology 21
A. Selecting Goals 21
B. Selecting Strategies 24
IV. Theory of Accountability: Representing Future and Current Generations 29
V. The Culture of Existential Advocacy 34
A. The Uncertainty Norm 34
B. The Deliberative Rationality Norm 35
C. The Supportive Dissent Norm 37
D. The Epistemic Identity Norm 39
VI. Discussion 42

INTRODUCTION

> "In some ways [the movement to mitigate existential risk] is the most inclusive movement, because it's trying to preserve everything. It sort of sits above all social movements…as a continuation of social justice. I think it's really vitally important to every person in the world today and every single future person…."
>
> "[When presenting existential risk to lawmakers, they often respond] 'How can I think about this when we've got so many crocodiles closer to the boat?'"
>
> "If the core aggrieved group is future generations, they don't get to chime in because: no time travel."
>
> \- Excerpts from research interviews

Imagine a pandemic 500 times more deadly than COVID-19, as scientists warn has been made more plausible with recent developments in synthetic biology.[1] Or imagine artificial intelligence so advanced that it threatens human control and even survival, a scenario that has moved from science fiction to a matter of serious academic and political concern with recent transformative advances in deep learning and a new wave of autonomous military applications.[2] Or imagine a war among the great powers that leads to massive detonation of nuclear weapons, a longstanding

---

[1] Note that COVID-19, according to a recent World Health Organization report, has been responsible for killing one out of 500 people globally. Hence, this article's reference to a pandemic "500 times more deadly than COVID-19" is meant to suggest an upper limit of mortality. *See* Thomas Mulier and Clara Hernanz Lizarraga*, Covid Killed One out of Every 500 People, WHO Report Shows*, BLOOMBERG (May 5, 2022), http://www.bloomberg.news/articles/2022-05-05/covid-killed-about-1-out-of-every-500-people-who-report-shows. For the scientific warnings that such extreme pandemics are growing more likely, *see* Kevin Esvelt, *How a Deliberate Pandemic Could Crush Societies and What to Do About It*, BULLETIN OF THE ATOMIC SCIENTISTS (November 15, 2022), https://thebulletin.org/2022/11/how-a-deliberate-pandemic-could-crush-societies-and-what-to-do-about-it/ (noting that advances in synthetic biology and "CRISPR-based gene drive systems" are increasingly able to produce new "pandemic-class agents" far more deadly than those found in nature, and describing scenarios where multiple synthetic pathogens could be released simultaneously); Abraar Karan and Stephen Luby, *A Natural Pandemic Has Been Terrible. A Synthetic One Would Be Even Worse*, STAT+ (Aug. 19, 2021), https://www.statnews.com/2021/08/19/natural-pandemic-terrible-synthetic-one-even-worse/ (describing pathogens that have been engineered to have the transmissibility of our most contagious diseases, e.g. measles, and the virulence of our most deadly diseases, e.g. Ebola); TOBY ORD, THE PRECIPICE: EXISTENTIAL RISK AND THE FUTURE OF HUMANITY 127-38 (2020) (detailing the history of gain-of-function research, including where researchers take highly fatal viruses and make them more contagious, and the historical record of laboratory leaks, information hazards, and biological warfare, which suggest multiple pathways for engineered pathogens to reach large populations). *See* discussion *infra* Part I.A.

[2] ORD, *supra* note 1 at 138-152 (noting survey findings suggesting that leading experts in artificial intelligence believe that the development of superintelligent systems is "plausible within a decade and more likely than not within a century," leading to a world where humanity cedes "our status as the most intelligent entities on Earth," and where it is difficult to align AI systems with human values and interests); Nick Bostrom, *Ethical Issues in Advanced Artificial Intelligence*, in COGNITIVE, EMOTIVE AND ETHICAL ASPECTS OF DECISION MAKING IN HUMANS AND IN ARTIFICIAL INTELLIGENCE (George Eric Lasker, Wendell Wallach, Iva Smit, eds., 2d ed. 2003) (detailing the challenges of engineering a safe "goal system" for superintelligent AI systems that would obey human commands without sacrificing human values and lives). *See* discussion *infra* Part I.A.

worry that has resurfaced with the invasion of Ukraine.[3] While the probability of such events occurring in the near-term may (hopefully) be quite low, researchers who study these risks have warned that these events seem to be growing more likely over time. Moreover, the magnitude of such catastrophes, were they to occur, would be so great that they may deserve serious attention even if the probabilities are very low.

An extraordinarily ambitious and well-funded community of advocates is focused on addressing these "existential risks," which they define as catastrophic events that would permanently bring an end to the meaningful existence of humanity.[4] This article offers the first empirical study of the lawyers and other legal advocates in this community. Drawing on a multi-method research design, I examine how social-change lawyering is being reimagined in the unique context of existential risk. What does it mean to be a lawyer for an issue, which, as described in the interviews excerpted above, operates on an unimaginably large scale (affecting "every person in the world and every single future person") but is obscured by a host of cognitive biases and political incentives directing our attention to issues that are seemingly more immediate (the "crocodiles closer to the boat")? Moreover, how can these lawyers faithfully represent future generations, a vast population of silent stakeholders in the decisions we make today (who cannot "chime in")? As detailed throughout this article, these advocates are engaged in a legal movement unlike any we have seen before, and with a distinct understanding of what it means to be a lawyer for social change.

Over the past two decades, existential risk has been the subject of a growing interdisciplinary academic field of inquiry. In the new canonical work of this field, *The Precipice: Existential Risk and the Future of Humanity* (2020), the philosopher Toby Ord estimated that the next century brings a one-in-six chance of existential catastrophe.[5] This estimate finds some support from other scholars, although always with the caveat that making such estimates is a highly speculative endeavor and could be off by multiple orders of magnitude in either direction.[6] Ord's

---

[3] ORD, *supra* note 1 at 90-102 (describing nuclear winter scenarios that might decimate the global food supply, and reviewing the Cold War history of occasions when full-scale nuclear attacks were nearly launched); Alan Robock, Luke Oman, and Georgiy L. Stenchikov, NUCLEAR WINTER REVISITED WITH A MODERN CLIMATE MODEL AND CURRENT NUCLEAR ARSENALS: STILL CATASTROPHIC CONSEQUENCES, 112.D13 J. GEOPHYSICAL RES. ATMOSPHERES (2007) (modeling nuclear winter scenarios based on the tools of climate science, which investigate how temperature changes would impact agriculture and the survival of humanity). *See* discussion *infra* Part I.A.

[4] *See generally,* Nick Bostrom, *Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards*, 9 J. EVOLUTION AND TECH. (2002) (introducing the notion of existential risk); ORD, *supra* 1 (defining existential threats as "risks that threaten the destruction of humanity's longterm potential," which is "most obvious" with human extinction but also includes other irreversible future trajectories of immense suffering). *See generally,* NICK BOSTROM AND MILAN M. CIRKOVIC, GLOBAL CATASTROPHIC RISKS; ANNETTE BAIER, RIGHTS OF PAST AND FUTURE PERSONS; Nick Bostrom, Existential Risk Prevention as a Global Priority; JOHN LESLIE, THE END OF THE WORLD: THE SCIENCE AND ETHICS OF HUMAN EXTINCTION.

[5] *See* ORD, *supra* note 1 (estimating a one in six chance ("Russian roulette") of existential catastrophe over the next century, and a one in three chance over the long term future).

[6] Anders Sandberg and Nick Bostrom, *Global Catastrophic Risks Survey*, https://www.global-catastrophic-risks.com/docs/2008-1.pdf (drawing on a survey of participants at the Global Catastrophic Risks Conference at Oxford University in 2008, finding an average estimate of 19% chance of human extinction prior to 2100, considering a number of risk categories from great power wars, pandemics and super-intelligent AI). Cf. WILLIAM MACASKILL, WHAT WE OWE THE FUTURE (2022) (suggesting a lower estimate of existential risk, likely below one percent over the next century). *See generally*, David Thorsadt, *Existential Risk Pessimism and Time of Perils* (2022) https://globalprioritiesinstitute.org/wp-content/uploads/David-Thorstad-Existential-risk-pessimism-.pdf; Michael

analysis takes into account scientific and theoretical investigations of a variety of threats, which he has investigated in his position as a senior fellow at Oxford University's Future of Humanity Institute. For members of Gen Z, this estimate of existential risk might not seem especially surprising, given their general sense that "humanity is doomed."[7] But this popular sense of doom is usually associated with climate change and other events that would devastate large populations over a number of generations, and thus are worthy of great concern, but are projected to be very unlikely to foreclose the long-term future of humanity. The greatest threats on the existential scale are thought to arise not from the familiar issues of climate change, asteroids, or nuclear weapons, but rather from new and emerging technologies, and the sense that we have only just started what will become a growing list of means to bring about existential catastrophes.[8] Even if one assumes relatively low probabilities to such events today—if this all sounds a bit too speculative and sci-fi at the moment—humanity will face this issue eventually. Assuming we continue to make scientific and technological progress, we can expect to produce new means of bringing about our own demise, including means that we cannot yet imagine.

The problem of existential risk is made more difficult by the global coordination that would be required to contain threats of this magnitude: An existential catastrophe arising in one jurisdiction would destroy humanity in all jurisdictions.[9] Thus, developing comprehensive regulatory responses in some but not all jurisdictions may entirely fall short. Moreover, by definition, we have no experience with events that terminate humanity. We cannot afford a single failure to prevent such a catastrophe, nor we can we rely on a reactive trial-and-error approach to designing our responses.[10] These observations have led some leading minds in this field to suggest that we are entering a new era of history, walking along a "precipice" in humanity's

---

Arid, *Database of Existential Risk Estimates*, EFFECTIVE ALTRUISM F. (Apr. 15, 2020), https://forum.effectivealtruism.org/posts/JQQAQrunyGGhzE23a/database-of-existential-risk-estimates.

[7] Caroline Hickman et al., *Young People's Voices on Climate Anxiety, Government Betrayal and Moral Injury: a Global Phenomenon*, https://www.sciencedirect.com/science/article/pii/S2542519621002783 (reporting a 10-country survey finding that a majority of people under 25 years old believe that "humanity is doomed."). *See* Angela Lashbrook, *'No Point in Anything Else': Gen Z Members Flock to Climate Careers*, GUARDIAN (Sept. 6, 2020), https://www.theguardian.com/environment/2021/sep/06/gen-z-climate-change-careers-jobs (describing the overwhelming concern for climate change among Gen Z as reflected in their career aspirations and choices). *See generally*, Madhukar Pai, *Young Climate Justice Activists Are Fighting for Our Collective Survival*, FORBES (July 28, 2022), https://www.forbes.com/sites/madhukarpai/2022/07/28/young-climate-justice-activists-are-fighting-for-our-collective-survival.

[8] *See, e.g.*, ORD, *supra* note 1 (estimating the existential risk associated with both nuclear war and climate change at 1 in 1,000 over the next century, but offering far higher estimates for unaligned artificial intelligence (1 in 10), engineered pandemics (1 in 30), and unforeseen anthropogenic risks (1 in 30)). *See also* Holden Karnofsky, Forecasting *Transformative AI, Part 1: What Kind of AI?*, COLD TAKES (Aug. 10, 2021), https://www.cold-takes.com/transformative-ai-timelines-part-1-of-4-what-kind-of-ai (suggesting that transformative artificial intelligence may radically accelerate scientific and technological development, which could lead to "technology capable of wiping humans out of existence").

[9] ORD, *supra* note 1 at (observing that the absence of strong multilateral mechanisms of global governance presents a formidable challenge for regulating these risks in the nearly 200 countries of the world).

[10] Nick Bostrom, *supra* note 1 ("Our approach to existential risks cannot be one of trial-and-error. There is no opportunity to learn from errors. The reactive approach—see what happens, limit damages, and learn from experience— is unworkable. Rather, we must take a proactive approach. This requires *foresight* to anticipate new types of threats and a willingness to take decisive *preventive action* and to bear the costs (moral and economic) of such actions.").

"most important century" as we make our initial encounter with existential threats.[11] As the issue is often framed in this field, humanity could either be approaching its end or, if we are able to survive the advent of technologies capable of producing existential catastrophes, and if we are able to establish global regulatory systems to assure that any new existential threats that arise would be contained as well, we could still be in the early infancy of humanity with an extraordinarily long arc of (hopefully good) experience ahead of us. This article contributes a new layer to this aspirational vision by considering what forms of legal advocacy might help in the process of establishing effective regulations.

The Oxford-based academic study of existential risk has now sparked a larger movement that not only conducts research but is increasingly working to intervene in existential threats. Non-profit organizations have proliferated in this field,[12] as well a new wave of student-led "existential risks initiatives" at leading universities, including Stanford, Cambridge, Harvard, and MIT.[13] The Stanford Existential Risks Initiative recently reported over 1,000 attendees from over 50 countries at the second meeting of their annual conference.[14] The field has seen extraordinary fundraising in recent years, currently totaling at least several billion US dollars, surpassing many of the leading philanthropic foundations of the world.[15] In just the past few years, this field has turned to the legal and political advocacy that is the subject of this article. This includes lobbying for legislation to prevent pandemics, threats from artificial intelligence, and global catastrophes more generally. It also includes efforts to encourage members of the existential risk community to run for U.S. Congress and other public office.[16]

---

[11] ORD, *supra* note 1; Holden Karnovsky, *The Most Important Century Blog Post Series, https://www.cold-takes.com/most-important-century/*. Cf., William MacAskill, *Are We Living at the Hinge of History?* 2–4 (Glob. Priorities Inst., Working Paper No. 12-2020), https://globalprioritiesinstitute.org/wp-content/uploads/William-MacAskill_Are-we-living-at-the-hinge-of-history.pdf (weighing arguments that we are unlikely to be living in the most important century).

[12] Organizations established around Oxford University include the Future of Humanity Institute, the Global Priorities Institute, and the Centre for Effective Altruism. Nearby Cambridge University is home to the Centre for the Study of Existential Risk. In the U.S., organizations working on existential risk include the Future of Life Institute and the Global Catastrophic Risk Institute.

[13] These student-led initiatives involve a range of activities including summer fellowships, conferences, and reading groups. *See generally* STAN. EXISTENTIAL RISK INITIATIVE, https://seri.stanford.edu (last visited Jan. 17, 2023); HARV.-MIT X-RISK, https://harvardmitxrisk.org (last visited Jan. 17, 2023).

[14] *See Stanford Existential Risks Initiative (SERI)*, STANFORD CENTER FOR INTERNATIONAL SECURITY AND COOPERATION: FREEMAN SPOGLI INSTITUTE, https://cisac.fsi.stanford.edu/stanford-existential-risks-initiative.

[15] *See* Naina Bajekal, *Want to Do More Good? This Movement Might Have the Answer*, TIME (Aug. 10, 2022, 7:00 AM), https://time.com/6204627/effective-altruism-longtermism-william-macaskill-interview (noting that, as of August 2022, the funding committed to Effective Altruism far exceeds that raised by the Ford Foundation (roughly $16 billion) and the Rockefeller Foundation (roughly $6 billion)). Funding committed to Effective Altruist causes has declined precipitously in recent months with the downfall of FTX and the decline in the net worth of other funders (most notably Dustin Moskovitz). Total funding committed in the field still likely exceeds $10 billion. *See* Benjamin Todd, *Is Effective Altruism Growing? An Update on the Stock of Funding vs People*, 80,000 Hours (July 28, 2021), https://80000hours.org/2021/07/effective-altruism-growing (noting the billions of dollars committed to Effective Altruism by Moskovitz and FTX as of July 2021); *see also* Benjamin Todd (@ben_j_todd), TWITTER (Aug. 20, 2022, 3:19 PM), https://twitter.com/ben_j_todd/status/1561100678654672896.

[16] *See, e.g.,* the campaign of Carrick Flynn for Oregon's Sixth District in the Democratic primary of 2022. Ian Ward, *The Esoteric Social Movement Behinds this Cycle's Most Expensive House Race*, Politico (May 12, 2022), https://www.politico.com/news/magazine/2022/05/12/carrick-flynn-save-world-congress-00031959.

This movement has only recently started exploring the role of law and lawyers. There are some initial signs that the legal profession might be receptive to addressing existential risk. The two most cited scholars in U.S. legal academia have written books on large-scale catastrophes, although neither connect their work to the Oxford-based literature on existential threats.[17] An international survey found that legal scholars generally believe that legal action taken today can impact existential risk and the long-term future.[18]

This optimism may be well placed, as the participants in this study have identified a new landscape of promising legal interventions, spanning from local efforts to prevent specific threats (e.g., seeking an injunction to prevent a laboratory from conducting gain-of-function research) to far-reaching efforts to establish the legal interests of future generations (e.g., judicial recognition of personhood, standing, or a right to life). As researchers in this field have reported, future generations are now referenced in 81 national constitutions,[19] in addition to a host of domestic law and international agreements,[20] and some courts have recently shown an unprecedented willingness to enforce these provisions in the context of climate change litigation.[21] Perhaps the most striking example was the 2021 German Constitutional Court decision striking down a national climate law while citing "intertemporal guarantees of freedom" and a "special duty of care…for the benefit of future generations."[22]

There is an emerging global wave of new legislation and parliamentary groups with a focus on protecting future generations from the impacts of low-probability/high-impact catastrophes.[23] In the U.S., new legislation (December 2022) requires federal agencies to assess and mitigate "existential risk," which is defined in the Act as risk with the "potential for an outcome that

---

[17] *See* RICHARD POSNER, CATASTROPHE: RISK AND RESPONSE (2004); CASS SUNSTEIN, WORST CASE SCENARIOS (2007); CASS SUNSTEIN, AVERTING CATASTROPHE (2021). *See also*, CHRISTOPH WINTER ET AL., LEGAL PRIORITIES RESEARCH: A RESEARCH AGENDA 5, 33-34 (2021), https://www.legalpriorities.org/research_agenda.pdf (observing that legal scholars have paid little attention to existential risk).

[18] *See* Eric Martinez and Christoph Winter, *Protecting Future Generations: A Global Survey of Legal Academics* (forthcoming).

[19] Renan Araújo and Leonie Koessler, *The Rise of the Constitutional Protection of Future Generations* (forthcoming).

[20] *Id.* (referencing the United Nations Charter's concern for "generations to come," the 1972 Stockholm Declaration concern for future generations and the environment, the 1992 Rio Declaration on sustainability into the future, and the 1997 UNESCO declaration on the rights of future generations).

[21] *See* Neubauer v. Germany, Bundesverfassungsgericht [BVerfGE] [Federal Constitutional Court of Germany] [GCC], 1 BvR 2656/18, 78/20, 96/20, and 288/20, para. 266, Mar. 24, 2020 (....); *see also* State of the Netherlands v. Urgenda, HR 20 december 2019, HAZA C/09/00456689, Urgenda/State of the Netherlands) (requiring emissions reductions on the grounds of right to life and right to respect for private and family life from Articles 2 and 8 of the European Convention on Human Rights and well as the UNFCC provision to "protect the climate system for the benefit of present and future generations of humankind"); Hague District Court 2021 (applying a "standard of care" as a human rights obligation that applies to future generations). *Cf.* Juliana v. United States, 947 F.3d 1159, 1175 (9th Cir. 2020) (finding that a group of twenty-one young plaintiffs lacked Article III standing for lack of redressability and dismissing all claims).

[22] *Id.*

[23] *See, e.g.*, ALL-PARTY PARLIAMENTARY GROUP FOR FUTURE GENERATIONS, https://www.appgfuturegenerations.com/about (aiming to establish "national wellbeing goals" relating to future generations); *Wellbeing of Future Generations Bill [HL]*, UK PARLIAMENT, https://bills.parliament.uk/bills/2869 (last updated May….2022). Colum Lynch, *DevExplains: In Short, Longtermism Has Arrived*, DEVEX (Dec. 22, 2022), https://www.devex.com/news/devexplains-in-short-longtermism-has-arrived-104626 (referencing institutions to protect future generations within the governments of Singapore, Finland, the United Arab Emirates, and Sweden).

would result in human extinction."[24] The Secretary General of the United Nations has issued a major report on existential risk and related threats to future generations. The UN has scheduled a "Summit of the Future" for 2024, which is expected to bring together heads of state from around the world to establish within the UN a Declaration on Future Generations, a Special Envoy for Future Generations, and an obligation to issue regular reports on global catastrophic risks.[25] The field of existential advocates examined in this article has taken a leading role in many of these and related efforts, although in order to preserve confidentiality I will only share details about specific advocacy projects where I have permission from research participants.

At the same time that these advocates have made such remarkable progress, they also been the subject of widespread backlash. Critical reactions to the movement intensified in recent months following the downfall of the field's top donor —Sam Bankman-Fried, who had pledged nearly all of his wealth, once valued at $29 billion, to the mitigation of existential risk but is now facing criminal and SEC fraud charges.[26] The common narrative of recent mass and social media suggests that advocating for the mitigation of existential risk tends to (perhaps by design) distract from the injustice and suffering that already exists in the world, instead directing attention toward the techno-utopian preferences of billionaire donors.[27] But these public reactions, whatever the ultimate merit of the critiques being raised, are based on a very shallow and often inaccurate view into the existential risk community, and generally no view at all into the field of legal and political advocates. This article provides an empirical window into this community drawing from ethnography, semi-structured interviews (n=53), and a systematic review of online materials.

The participants in this study included members and affiliates of over two dozen organizations,[28] but my main focus is the Legal Priorities Project (hereinafter "LPP"). I will profile LPP throughout this article. As background for this discussion, I offer a brief description of the organization here. Founded in 2020 as a 501(c)(3) by students and a visiting professor at Harvard Law School, LPP is the only legal organization in the world focused on existential risk. In its first year of operation, LPP wrote a research agenda and began to build a network of young lawyers and law students via summer fellowships and speaker programs. In their second year,

---

[24] The Global Catastrophic Risk Management Act of 2022, H.R. 7776, 117 Cong. (2022).

[25] *See* ANTÓNIO GUTERRES, UNITED NATIONS, OUR COMMON AGENDA: REPORT OF THE SECRETARY-GENERAL 27, 64 (2021), https://www.un.org/en/content/common-agenda-report/assets/pdf/Common_Agenda_Report_English.pdf (referencing "solidarity" between current and future generations); *See also* the United Nations Development Programme (UNDP) 2020 Human Development Report, which included an essay on existential risk written by Toby Ord).

[26] Gerrit De Vynck, *U.S. Charges FTX Founder Sam Bankman-Fried with Criminal Fraud*, WASH. POST (Dec. 13, 2022, 2:57 PM), https://www.washingtonpost.com/technology/2022/12/13/sbf-sec-fraud-charges.

[27] *See* Jennifer Szalai, *How Sam Bankman-Fried Put Effective Altruism on the Defensive*, N.Y. TIMES (Dec. 13, 2022), https://www.nytimes.com/2022/12/09/books/review/effective-altruism-sam-bankman-fried-crypto.html (citing the keys texts on existential risk for the general view that "considerations of immediate need pale next to speculations about existential risk — not just earthly concerns about climate change and pandemics but also…more extravagant theorizing about space colonization and A.I."); Emily Frey & Noah Glanslracusa, *The Moral Failing of Effective Altruism*, BOS. GLOBE (Nov. 22, 2022, 3:00 AM), https://www.bostonglobe.com/2022/11/22/opinion/moral-failing-effective-altruism (suggesting that tech billionaires may tend to be especially drawn to the notion that emerging tech poses a major threat to the future of humanity because this issue is framed as matter relating to their technological expertise).

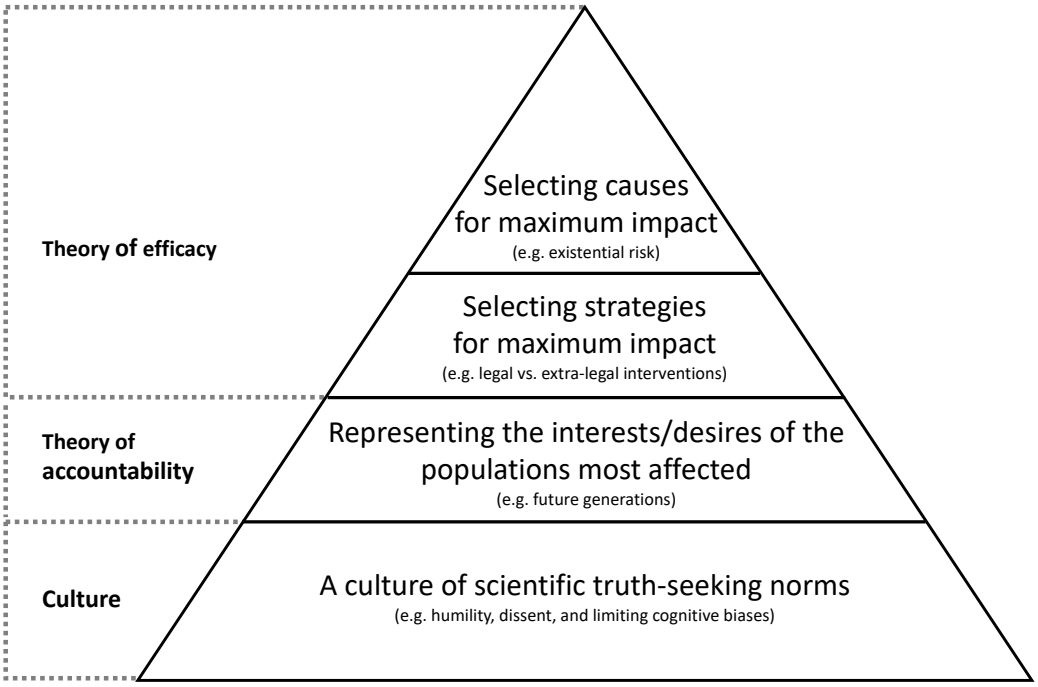
[28] In total interview participants were employed full-time by 18 different organizations, while some participants had multiple affiliations and several were occupied with studies at the undergraduate or graduate level.

they shifted attention to policy advising, providing a legal perspective on new legislative efforts relating to existential risk. They also recently formed an impact-litigation team, hiring a Costa Rican human rights attorney, a Dutch judge, and an American public interest lawyer. They are a highly global organization—the membership has a clear tilt toward Europe and the U.S., but, over the period of ethnography, the roughly twelve members who attended LPP's internal all-hands meetings included nationalities (and legal training) from Africa, Australia, Central America, Continental Europe, North America, and South America.[29] They operate remotely with occasional in-person meetings. Thus, the study was largely conducted via video-conferencing calls in addition to in-person visits to key sites relating to LPP and other advocacy organizations in e.g., Oxford, London, Geneva, Washington DC, San Francisco, and Boston.

The primary empirical finding of this article is that these advocates have developed a distinct model of social-change lawyering—the "priorities methodology." In short, this model begins with first-principle moral commitments, e.g. enhancing overall human well-being, reducing extreme suffering, or promoting justice. They then apply a set of criteria (e.g. importance, neglect, and tractability) to choose causes where they believe they can have the greatest impact on their moral commitments. Existential risk is prioritized by these advocates because it is a neglected issue that could severely affect a great number of people. But a willingness to change priorities upon new evidence is a key commitment under this model. Advocates then analyze strategic decisions in an effort to determine which option has the greatest expected value toward their chosen goals. As observed in this study, this model's commitment to optimizing impact is supported by formal decision-making processes as well as a daily culture of reinforcing scientific, truth-seeking norms, including (1) embracing epistemic humility and normalizing uncertainty, (2) fostering a warm, inclusive, and empathetic atmosphere of "supportive dissent," and (3) limiting cognitive biases that would interfere with their focus on maximizing impact. In practice, these advocates are remarkably adherent to this set of cultural norms, although they acknowledge that these norms conflict with some aspects of human nature and professional identities as lawyers. Moreover, this model raises difficult points of tension when seeking to represent the interests of voiceless future generations while also working to incorporate a broader set of current-person voices in strategic decision-making processes. The full model is depicted in Figure 1.

[29] It should be noted that this degree of geographic diversity is perhaps uncommon among organizations in this field, although LPP is not alone in seeking to diversify the geography, race, gender and other identities represented in the movement. This issue is discussed infra Part IV.

Figure 1. The Priorities Methodology for Social-Change Lawyering



In some respects, the priorities methodology is consistent with leading recommendations from the literature on law and social change. Empirical research has revealed that the majority of public interest lawyers lack explicit theories of change, formal processes for setting priorities, and measures of their performance.[30] Scholars have called on social-change lawyers to adopt more rigorous approaches to strategic planning and assessment with a greater focus on de facto rather than merely de jure impact.[31] The existential advocates seem to take up this call very directly, even taking it to an extreme of sorts with their commitment to maximizing their overall counterfactual impact.[32]

But these advocates depart from the literature's recommendations in one key respect: While they recognize that favorable public opinion and grassroots mobilization have supported the efficacy and accountability of legal activism in other contexts, these advocates worry that "going broad" would tend to compromise their priorities methodology and politicize the issue of existential risk. The cultural values outlined in this article, e.g. seeking to limit identity, emotions, groupthink, and claims of certainty, are seemingly point-by-point the exact opposite of the "mobilizing frames" that scholars have found to be the essential ingredients of successful broad outreach campaigns and movement building. This observation reveals the central point of tension explored throughout this article. These advocates seek to enhance their inclusiveness and

[30] *See generally* Deborah L. Rhode, *Public Interest Law: The Movement at Midlife*, 60 STAN. L. REV. 2027 (2008). (finding that leaders of public interest law organizations generally lacked "explicitly articulated theories" to select priorities and strategies and did not have "specific measures of performance," with only 14% undertaking "extensive" formal decision-making processes).
[31] Id.
[32] Cummings forthcoming…

democratic legitimacy without undermining what their community does so well—overcoming cognitive biases to maintain a focus on the protection of future generations from neglected low-probability/high-impact events.

Before detailing these empirical findings, Part I provides background about (1) the issue of existential risk and (2) the socio-legal literature on the question of how law and lawyers contribute to movements for social change. Part II describes the methods of this study. Parts III through V present the empirical description of existential advocacy and the priorities methodology, beginning with theories of efficacy and accountability, and then discussing the movement's underlying cultural commitments. Part VI concludes the article with an assessment of existential advocacy, as it is currently practiced, and makes recommendations for adapting this approach as this movement scales up and pursues more direct and high-profile legal interventions.

I. Background:

A. Existential Risk

On the scale of human history, it is only very recently that we have developed technologies capable of foreclosing the human future.[33] Professor Nick Bostrom, the Oxford philosopher who coined the term "existential risk" to refer to threats of human extinction or otherwise irreversible destruction of human potential, notes that until the advent of nuclear weapons there were "probably no significant existential risks in human history…and certainly none that it was within our power to do something about."[34] Even nuclear weapons are often categorized as sub-existential threats, although they could theoretically end humanity via a "nuclear winter" if deployed ubiquitously around the globe.[35] Researchers in the emerging field of existential risk studies have now spent over two decades analyzing a wide range of conceivable threats, which has led many in this field to the conclusion that existential risk should be assigned a substantial probability.[36] Ord suggests we are entering a new historical era, which he labels "the Precipice," defined by our initial meeting with threats of this magnitude.[37] Ord's optimistic framing of this

---

[33] *See* THOMAS MOYNIHAN, X-RISK: HOW HUMANITY DISCOVERED ITS OWN EXTINCTION (2020) (discussing the long-running history of religious traditions of apocalyptic prophecy, which portend a "sense of an ending," but explaining that existential risk, and human extinction in particular, is a "comparatively novel idea" involving an "ending of sense," a more complete ending of human experience).

[34] Nick Bostrom, *Dinosaurs, Dodos, Humans?*, 2006 GLOB. AGENDA 230, 230.

[35] *See* Alexandra Witze, *How a Small Nuclear War Would Transform the Entire Planet*, NATURE (Mar. 16, 2020), https://www.nature.com/articles/d41586-020-00794-y.

[36] *See* Bostrom, *supra* note (defining existential risks as threats to "either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential"); ORD, *supra* note 1 (defining existential risk to include extinction, locked-in totalitarianism ("world in chains"), and irreversible collapse of civilization "where humanity across the globe loses civilization, a world without writing, cities, law, or any trappings of civilization."). *See generally, Policy Idea Database*, CNTR. STUDY EXISTENTIAL RISK, https://www.gcrpolicy.com/ideas (last visited Aug. 15, 2022) (showing a total of 376 publications in the field of existential and catastrophic risk studies with an increase after 2017 to 75 to 82 publications per year); *see also* Rumtin Sepasspour, *Release of Existential Risk Policy Database*, EFFECTIVE ALTRUISM F. (Feb. 7, 2022), https://forum.effectivealtruism.org/posts/5GnFLDfnzmK3Gv4gB/release-of-existential-risk-policy-database.

[37] ORD, *supra* note 1.

era is that it will (hopefully) be remembered as the time when humanity "opened its eyes" to existential risk and "guaranteed a long and flourishing future" through transformations in our legal, political, normative, and cognitive frameworks.[38]

Ord offers a set of best-guess estimates of how likely it is that we would experience different existential events. While conceding that assigning these probabilities is highly speculative, he notes that it can be a helpful exercise in comparing different threats and avoiding the human tendency to dismiss risks that are vaguely described as "unlikely." While climate change and nuclear weapons are familiar issues of public concern, Ord assigns these categories only a 1/1000 chance of reaching the scale of irreversibly destroying the human future over the next century.[39] Existential threats from natural sources are similarly considered quite unlikely to materialize in the near future. Although asteroids loom large in fictional accounts, as well in prehistory as the source of the Cretaceous–Paleogene Extinction, this category is now well studied.[40] By looking to the skies and to the earth (via the geological record), scientists have shown that such events occur infrequently and thus appear to pose a relatively slight danger in the near term.[41] Ord suggests a dramatic contrast between the probability of natural and anthropogenic existential risks. For example, he estimates a 1/10,000 probability that natural pandemics will bring about an existential catastrophe over the next century, while engineered pandemics are assigned a probability of 1/30.[42] Moreover, engineered pandemics are part of a growing category of concerns relating to synthetic biology, including bioweapons, pathogens escaping laboratories by accident or on purpose, or information hazards (where information required to create dangerous biological materials are published widely or otherwise not kept confidential), and under-regulation of the burgeoning DNA synthesis industry. More speculatively, other frontiers of scientific endeavor could also pose existential threats, such as bringing back unpredictable materials from space exploration or running "radical scientific experiments" with unknown risks that could theoretically reach the scale of destroying life on Earth and beyond.[43]

But for Ord, along with most scholars in this field, the greatest single category of existential risk relates to the development of transformative artificial intelligence ("AI"), estimated by Ord as a 1/10 existential threat over the next century. With new developments in deep learning, a growing number of AI experts now believe that it will be possible to develop AI systems that far surpass

---

[38] *Id*., at 31.

[39] ORD, *supra* note 1 (noting that nuclear weapons and climate change "awoke us to the possibilities of destroying ourselves," and noting that the threat of climate change could grow to existential proportions with a runaway greenhouse effect that boils the oceans or sets off a cascade of ecosystem failures, although experts tend to suggest that such total destruction scenarios are unlikely). Existential risk is often viewed as a subset of the broader category of global catastrophic risk, which threaten widespread and even global harm but may lack the "terminal" intensity of risks on the existential scale. *See* BOSTROM AND CIRKOVIC*, supra* note (defining global catastrophic risk as "a catastrophe that [would cause] 10 million fatalities or 10 trillion dollars worth of economic loss").

[40] Adam Mann, *Odds of Death by Asteroid? Lower Than Plan Crash, Higher Than Lightning*, WIRED (Feb. 15, 2013, 2:39 PM), https://www.wired.com/2013/02/asteroid-odds (reporting on the risk table by NASA's Near Earth Object program that reports on the "likelihood of impact for the next 100 years" of known asteroids).

[41] BOSTROM AND CIRCOVIC, supra note.

[42] *See* ORD, *supra* note 1 (reviewing the long history of the use of disease as a weapon, as well as examples of lab escapes and information hazards including the publication of the small pox genome).

[43] *Id.* (discussing "radical science experiments," and citing the example of the advent of nuclear weapons when some scientists theorized that the first detonation would ignite Earth's atmosphere and set off an existential catastrophe).

human intelligence in many or all respects. There is less consensus about when such technologies would be developed, although many experts believe that it is only a matter of decades or sooner.[44] These timelines have grown shorter with the advent of generative AI—systems that write essays, poetry, and music, create original images and video, and conduct scientific inquiries, as well as systems that consistently beat humans at games based on psychology and stratagem (e.g. poker and Diplomacy).[45]

Advanced AI systems could give rise to existential risks if deployed maliciously by bad actors (i.e. with autonomous weaponry). But a perhaps more fundamental concern is that such systems might be made in an inherently unsafe manner by actors who are in a race to develop the most powerful technologies and deploy them quickly and expansively.[46] This race dynamic appears to be emerging not only in the private sector but also among national governments who see enormous economic and military value in this technology, as described in the Chinese political leadership's 2030 goal to achieve "AI dominance," and recent US efforts to limit Chinese access to high-end semiconductors.[47] An even more fundamental issue is that it may be difficult or impossible, even with our best efforts, to align these systems with human values and interests—that is, to design neural networks or reinforcement learning systems with reward functions so that they reliably respect, at a minimum, human life and our interest in avoiding locked-in dystopic futures. One of the concerns arising around efforts to reduce algorithmic bias (e.g. AI outputs that demonstrate bias on the grounds of race or gender) is that this may be a sign of the more general challenge of aligning transformative AI systems with good human values—moreover, this challenges is exacerbated by the further question of how to determine what good human values are and who should get to decide this.[48] Ord and others also note that transformative AI may have both positive and negative effects on other existential threats, e.g. helping us develop tools to deal with problems like climate change but rapidly accelerating the scientific advances that give rise to new threats on the existential scale.

Finally, the notion of "unforeseen" risks may seem hopelessly vague and speculative. But there may be reason for special concern regarding this category. In his writing on the "vulnerable world hypothesis,"[49] Bostrom posits that humanity's never-ending practice of drawing out new inventions from the metaphorical "urn of creativity" may eventually yield a "black ball" technology, that is, a technology that "invariably or by default destroys the civilization that

---

[44] *See*, *Data of Artificial Intelligence*, https://www.metaculus.com/questions/5121/date-of-artificial-general-intelligence/ (reporting on a monthly basis the average expert estimate of the arrival of artificial general intelligence, which has wavered in recent months between the years 2034 and 2059, but noting that roughly 50% of AI researchers believe that such technology will not arrive until after the year 2100).

[45] Seb Krier, *AI from Superintelligence to ChatGPT* (2022) https://www.worksinprogress.co/issue/ai-from-superintelligence-to-chatgpt/?fbclid=IwAR2_TBKaWCxIwF0XEPB1MEcs9DPHN8bcl0xBlq--DDDc53g2wHLR579D7T8.

[46] *See generally* Wim Naudé & Nicola Dimitri, *The Race For An Artificial General Intelligence: Implications for Public Policy*, 35 AI & SOC'Y 367, 367 (2020), https://doi.org/10.1007/s00146-019-00887-x.

[47] Kathleen Walch, *Why the Race for AI Dominance is More Global Than You Think*, FORBES (Feb. 9, 2020, 1:00 AM), https://www.forbes.com/sites/cognitiveworld/2020/02/09/why-the-race-for-ai-dominance-is-more-global-than-you-think.

[48] *See generally*, Iason Gabriel, *Artificial Intelligence, Values, and Alignment*, 30.3 MINDS AND MACHINES 411 (2020); NICK BOSTROM, SUPERINTELLIGENCE (2014).

[49] *See generally* Nick Bostrom, *The Vulnerable World Hypothesis*, 10 GLOB. POL'Y 455, 455 (2019), https://doi.org/10.1111/1758-5899.12718.

invents it."[50] An example could be a device with the power of nuclear weapons but an ease of assembly that requires only a commercially available 3D printer or the equipment in a typical garage.[51] It is perhaps only by good fortune that humanity has not yet discovered a black ball, and it may be only a matter of time before we do. If we create transformative AI that greatly accelerates scientific and technological development (what is sometimes called in the existential risk community, "PASTA," a "Process for Automating Scientific and Technological Advancement"), the invention of black-ball technologies may grow increasingly likely.[52]

Estimates of existential risk are based not only on assessing the destructive capacities of various technologies but also, crucially, on assessing our ability to prevent catastrophic use of those technologies. When Ord offers 1/6 estimate of existential risk this century he notes that this estimate already assumes that we would cut existential risk by half, because we would, in the coming decades, "get our act together and start taking these risks very seriously."[53] Any such awakening in our political and legal systems is inhibited by a wide range of cognitive biases that make it difficult to recognize the scope of existential risk. To cite just a few key examples, people generally find it difficult to imagine events on a scale we have never seen before (the availability heuristic) and with a moral lens that is evolutionarily tuned to small-scale and nearby harms (scope neglect) that befall known individuals (the identifiable victim effect) who are alive today (present bias).[54] We tend to dismiss threats of low probability events unless our emotions are primed.[55] In popular understandings of existential risk, these cognitive biases may be exacerbated by the association of existential risk with science fiction, irrational doomsayers and "preppers," and, in some emerging caricatures of the current existential risk community, hyper-rational "tech bros."[56] In the political realm, these cognitive biases may be exacerbated by short-

---

[50] *Id.*

[51] Id. at 455-56 (discussing the possibility of apocalyptic technology created "with a piece of glass, a metal object, and a batter arranged in a particular configuration").

[52] Holden Karnofsky, *Forecasting Transformative AI, Part 1: What Kind of AI?*, COLD TAKES (Aug. 10, 2021), https://www.cold-takes.com/transformative-ai-timelines-part-1-of-4-what-kind-of-ai.

[53] ORD, *supra* note 1.

[54] *See* Tyler M. John & William MacAskill, *Longtermist Institutional Reform* 5 (Glob. Priorities Inst., Working Paper No. 14-2020), https://globalprioritiesinstitute.org/wp-content/uploads/Tyler-M-John-and-William-MacAskill_Longtermist-institutional-reform.pdf. (discussing cognitive biases relating to future generations); CASS R. SUNSTEIN, BEHAVIORAL SCIENCE AND PUBLIC POLICY 3 (2020) (noting that our concern is generally diminished when catastrophic threats are framed as a risk primarily to future generations—those living in "Laterland"); Winter et al, *supra* note (noting that human cognition tends to be limited in its ability to consider "the vastness of the future, in particular…human extinction scenarios.").

[55] CASS SUNSTEIN, WORST CASE SCENARIOS (describing the general lack of emotional response around threats that are rare or unprecedented, and our tendency to over-react when low-probability events are emotionally salient, such as when such events have occurred recently); CASS SUNSTEIN, AVERTING CATASTROPHE (observing that some catastrophes are the result of exponential rather than linear growth, which leads to under-reaction due to "exponential growth neglect"); Eliezer Yudkowsky, *Cognitive Biases Potentially Affecting Judgement of Global Risks*, *in* GLOBAL CATASTROPHIC RISKS 91, 91–115 (Nick Bostrom & Milan M. Cirkovic, eds., 2d ed. 2020) (summarizing the relevant cognitive biases and noting that humanity tends to "overestimate the predictability of the past and underestimate the surprise of the future").

[56] *See* Joshua Schuster and Derek Woods, *Calamity Theory: Three Critiques of Existential Risk*, https://manifold.umn.edu/read/calamity-theory/section/3a175630-820b-4886-aadc-06f3b4021516 (expressing a general wariness in relation to the observation that "existential risk theory has been conducive to a warm reception by a 'tech bro' Silicon Valley audience."); POSNER, *supra* note (arguing that irresponsible doomsday predictions can lead to a backlash of excessive optimism regarding the risk of large-scale catastrophe, but noting that thoughtful science fiction can also be helpful to illuminate these threats).

term incentives that prioritize currently living (and voting and lobbying) people over future generations.[57] Moreover, any effort to mitigate existential must overcome collective action problems associated with goods that are public, global, and intergenerational.[58] These political conditions could be made far more perilous by "existential risk factors" in the coming years, such as great power wars, extreme environmental impacts, or other events that make it less likely that we, as a global community, will be willing and able to work together to address threats on the existential scale.

But maybe these "biases" against concerning ourselves with existential risk are actually pointing toward something true. Should we care about preventing existential risk? Does existential risk matter? This question has both empirical and normative dimensions. This section has so far focused on empirics—assessing the probability that an existential event will occur. If this probability is extremely low (or extremely high and intractable), we might have little reason to invest in the prevention of existential threats. The normative dimension asks whether an existential event would be undesirable. The simplest response from participants in this study is that existential risks threaten to harm a great number of people who actually exist. This includes people whose lives would be ended (perhaps involving immense suffering) by an existential catastrophe or whose lives would be made much worse under an unrecoverable dystopia scenario. Moreover, efforts to mitigate existential risk greatly overlap with efforts to mitigate sub-existential risk—and these smaller catastrophes may be more likely to occur in the near-term and would have devastating effects on current-living people.

The more complex normative response considers the impact of human extinction on future generations, where the impact would be "felt" by people who would not have the opportunity to exist. A full treatment of this issue and related theories of population ethics is beyond the scope of this article.[59] This issue is the subject of a new field of philosophical inquiry known as "longtermism," which considers the moral weight of future generations.[60] But it is worth noting briefly that many participants in this study analyze this question in the following terms: potential future persons could outnumber us (current persons) to such a radical degree that if one assigns any non-negligible value to the existence of future persons, and if one believes that future persons will tend to have lives worth living, one might view human extinction as a great harm.[61] Combining these concerns relating to human extinction scenarios (for both the actual people who would die in such a catastrophe and the potential future people who would not have a chance to

---

[57] Tyler M. John, *Representing Future Generations*, YOUTUBE (Mar. 21, 2020), https://youtu.be/095kFEA-jpE (observing that elected officials and other political leaders tend to consider future effects only on the scale of 2-5 years or "the next decade," due to cognitive biases, time preference, and election incentives).
[58] *See* ORD, *supra* note 1.
[59] *See generally* DEREK PARFIT, REASONS AND PERSONS (2d ed., 1986) (introducing population ethics with the hypothetical comparison of an event that ends 100% of human life and an event that ends 99% of human life, where the latter event permits the continuation of future generations, thus raising the question of how much we should value people who could one day exist).
[60] *See generally,* Bostrom, *supra* note; Hilary Greaves and William MacAskill, *The Case for Strong Longtermism* (2019), https://globalprioritiesinstitute.org/hilary-greaves-william-macaskill-the-case-for-strong-longtermism-2/. Bostrom (x-risk as global priority); Greaves and MacAskill 2019; John 2020, Tarsney 2019…MacAskill 2022.
[61] Holden Karnofsky, *Debating Myself on Whether "Extra Lives Lived" Are As Good As Deaths Prevented*, COLD TAKES (Mar. 29, 2022), https://www.cold-takes.com/debating-myself-on-whether-extra-lives-lived-are-as-good-as-deaths-prevented.

exist) and permanent dystopic scenarios (for many actual people current and future) makes a strong case, as participants in this study argue, that existential risk matters.

B. Literature on Law and Social Change

As context for this article's discussion of social-change lawyering among the existential advocates, this section provides a brief background on how scholars have generally described the role of lawyers in movements for social change.[62] Much of this literature has focused on the dark side of movement lawyering, criticizing lawyers who exaggerate the value of court-led social change, especially where lawyers fall under a "myth of rights" and a "hollow hope" that de jure victories in the courts will, on their own, bring about de facto social change.[63] The lawyers of the Civil Rights Movement have been cited as examples of this overly legalistic orientation, prioritizing litigation while discouraging grassroots organizing and legislative advocacy.[64] These scholars point to examples where judicial victories have sparked backlash and countermobilization, undermining the impact that court-centered activism seemed to promise.[65] Moreover, some scholars note that lawyers tend to dominate movement agendas, marginalizing the grassroots voices of the most affected constituencies.[66] Lawyers may tend to de-radicalize movements, both because of the lawyers' own preferences for working within institutional channels and because of the nature of the law as a conservative, precedential system for maintaining the status quo of the legal order.[67] These observations have led to calls for lawyers to take a reduced role in movement leadership.[68]

---

[62] *See generally*, Sameer M. Ashar, "Movement Lawyers in the Fight for Immigrant Rights." UCLA L. REV. 64 (2017): 1464; TOMIKO BROWN-NAGIN, COURAGE TO DISSENT: ATLANTA AND THE LONG HISTORY OF THE CIVIL RIGHTS MOVEMENT (2011); Susan D. Carle, "A Social Movement History of Title VII Disparate Impact Analysis," FLA. L. REV. 63 (2011): 251; SCOTT CUMMINGS BLUE AND GREEN; AN EQUAL PLACE; MICHAEL MCCANN, RIGHTS AT WORK; MICHAEL J. KLARMAN, FROM JIM CROW TO CIVIL RIGHTS: THE SUPREME COURT AND THE STRUGGLE FOR RACIAL EQUALITY (2006); KENNETH W. MACK, REPRESENTING THE RACE; Douglas NeJaime, "Marriage Equality and the New Parenthood," 129 HARV. L. REV. 1185 (2015); SARAT AND SCHEINGOLD, CAUSE LAWYERS AND SOCIAL MOVEMENTS (2006); ANN SOUTHWORTH, LAWYERS OF THE RIGHTS: PROFESSIONALIZING THE CONSERVATIVE COALITION (2019).

[63] *See* GERALD N. ROSENBERG, THE HOLLOW HOPE: CAN COURTS BRING ABOUT SOCIAL CHANGE (2d ed., 2008); STUART A. SCHEINGOLD, THE POLITICS OF RIGHTS: LAWYERS, PUBLIC POLICY, AND POLITICAL CHANGE (2d ed., 2004).

[64] *See* MICHAEL J. KLARMAN, FROM JIM CROW TO CIVIL RIGHTS: THE SUPREME COURT AND THE STRUGGLE FOR RACIAL EQUALITY (2004).

[65] *See,* ROSENBERG, *supra* note.

[66] *See* Bell, *supra* note.

[67] *See* Catherine Albiston, *The Dark Side of Litigation as a Social Movement Strategy*, 96 IOWA L. REV. BULL. 61, 62 (2011) (observing that lawyers often "deradicalize and subtly reshape social movements"); Scott Cummings, *Law and Social Movements: An Interdisciplinary Analysis*, *in* HANDBOOK OF SOCIAL MOVEMENTS ACROSS DISCIPLINES 233, 263 (C. Roggeband & B. Klandermans, eds., 2017) (noting that legal framings can "sanitize" issues to "comport with mainstream values," perhaps because, as CLS scholars have long argued, law favors the status quo and legal victories do not transform structural relations); Scott L. Cummings & Deborah L. Rhode, *Public Interest Litigation: Insights From Theory and Practice*, 36 Fordham Urb. L. J. 603, 612 (2009) (noting that legal actions can dissipate grassroots activism, thereby reducing a "movement's transformative potential").

[68] SCOTT CUMMINGS, MOVEMENT LAWYERING forthcoming (describing the "movement liberalism" literature where scholars recommend that social-change lawyers take a more limited, conventional client-centered advocacy role in support of grassroots organizations).

In contrast, recent empirical studies challenge the portrayal of social-change lawyers as narrowly legalistic and strategically unsophisticated.[69] Scholars increasingly see a revival of movement lawyering under the rubric of "integrated advocacy," in which lawyers coordinate their distinctively legal work (litigation and other legal services) with other tactics such as building movements, shaping public opinion, and advocating for new legislation.[70] By embedding lawyers within movements, integrated advocacy serves to enhance lawyers' accountability to the populations most affected by an issue. It also appears to enhance efficacy in at least some contexts. Professor Scott Cummings has described in empirical detail several examples of this "new canon" of social-change lawyering.[71] He notes that the marriage equality movement in particular has reshaped the test case litigation model with a greater emphasis on fostering favorable public opinion ("hearts and minds") through local legislative campaigns.[72] This creates a sense of collective demand for reform, which may help to persuade judges to make what would have previously seemed very bold decisions (e.g., *Obergefell v. Hodges* 2015) while also persuading the public to support enforcement of those decisions.[73] Cummings stresses that social change is a long-term project, marked by continual dynamics of resistance and struggle.[74] Law, when coordinated with action in other strategic domains, can be a powerful tool in these ongoing struggles.

This literature on social-change lawyering has focused primarily on lawyers within progressive grassroots movements—generally where a community turns to law in an effort to find legal voice and remedies. Some scholars have extended this inquiry to other contexts, including movements for animals, the environment, and conservative causes, where the grassroots element is often less salient.[75] The existential advocates are a step further in this direction away from traditional understandings of social movements. If their primary constituency is future generations, it is not possible for the existential advocates to develop a grassroots movement where the most affected community would organize to form a collective voice.[76]

---

[69] *See* Alan K. Chen, *Rights Lawyer Essentialism and the Next Generation of Rights Critics,* 111 MICH. L. REV. 903, 905-06 (2013) (describing "rights lawyer essentialism" as a common but inaccurate portrayal of civil rights attorneys as "elitist, singularly minded litigation hawks who care little for their clients or the subtleties of the dialectic political process."); *see also* ALAN K. CHEN AND SCOTT L. CUMMINGS, PUBLIC INTEREST LAWYERING: A CONTEMPORARY PERSPECTIVE 518 (2013) (providing examples of how cause lawyers conceive of litigation "not in isolation, but as part of a comprehensive set of tools that are useful in advancing social reform.").

[70] SCOTT CUMMINGS, LAWYERS AND MOVEMENTS (forthcoming) (observing that movement lawyering is experiencing a "revival" after "decades of dormancy").

[71] Id.

[72] Id.

[73] Id. (observing that lawyers can and should contribute to all aspects of this integrated model, and can justifiably take leadership positions where they should be subject to the same scrutiny as other movement leaders).

[74] Id.

[75] *See* Tomiko Brown-Nagin, *Elites, Social Movements, and the Law: The Case of Affirmative Action*, 105 COLUM. L. REV. 1436, 1508 (2005) (noting that movements are generally made up of "socially marginal citizens" responding to oppression and inequality); Scott L. Cummings, *Law and Social Movements: Reimagining the Progressive Canon*, 2018 WIS. L. REV. 441, 451–60, 470–78, 487–94 (2018) (describing the "contemporary progressive legal canon" rooted in concern for marginalized groups, inequality, and a struggle over resources); David A. Snow, Sarah A. Soule, Hanspeter Kriesi, *Mapping the Terrain*, *in* THE BLACKWELL COMPANION TO SOCIAL MOVEMENTS 3–16 (2004); SOUTHWORTH, *supra* note.

[76] Toby Ord, Senior Rsch. Fellow, Oxford Univ & Author, *The Precipice*, Stanford Existential Risks Initiative Virtual Conference (Apr. 17, 2021) (noting that the population who "bear the most of the relevant costs" would usually vocalize their interest and "campaign and push for change.").

Yet, the existential advocates do resemble a social movement in some key respects. The effort to enfranchise a voicelessness population of future generations is similar to the classic concern for marginalized communities in the civil rights tradition. Moreover, some participants in this study very much view their activism as an extension of their past contributions to progressive social justice movements. For these advocates, reducing existential risk is framed as an effort to counteract discrimination against future persons and to advance equality, as one participant explained: "Our assumptions are quite simple. We want to treat everyone as equal, not just in space, but also in time."[77] Participants in this study are engaged with the same strategic questions debated throughout the literature on movement lawyering, e.g. how to enhance de facto change and how to best represent the preferences and interests of key constituencies. With these commonalities in mind, this article will refer to the community working on existential risk as a nascent "movement." This allows for an inquiry into whether, and to what extent, insights from socio-legal literature apply to the novel context of existential risk mitigation. It also allows for a strategic analysis of whether this movement should broaden into something more closely resembling a grassroots social movement.

II. Research Design

This study consisted of a two-year study including ethnography, interviews, analysis of online materials, and first-hand experience in the field. The ethnography was conducted at the Legal Priorities Project ("LPP") over a period of five months in 2021 and 2022, during which 59 virtual meetings were observed. Ethnography is an anthropological method of participant observation, in which the researcher is invited to join a community while taking detailed analytic and descriptive notes on norms, interactions, and other cultural dynamics, as well as the researcher's own experiences of membership in the community.[78] Rather than testing hypotheses, ethnographers tend to inductively generate "grounded theory," wherein the study's key theoretical observations emerge from the thematic coding and analysis of fieldnotes.[79] This requires deep immersion in the culture under study. LPP was highly receptive to this methodology. They permitted me to observe weekly all-hands and small-group meetings, research workshops, textual discussions, and shared documents.

Following this period of formal ethnography, I have continued to regularly engaged with LPP and the larger legal community working on existential risk. This has included visits to key sites of the movement (e.g., Oxford, London, Geneva, Washington DC, San Francisco, and Boston). I attended twelve conferences and workshops over this period, while regularly following online discussions (e.g., Slack workspaces, Facebook groups, blogs, Discord channels, podcasts, newsletters, and the Effective Altruism Forum). A range of published online materials are also referenced throughout this article, including white papers, research agendas, and curricula for courses and reading groups.

---

[77] One could also view existential risk mitigation in terms of a "struggle over resources" (Cummings on progressive moments, supra), with the competing parties here being current and future generations.

[78] *See* KAREN O'REILLY, ETHNOGRAPHIC METHODS (2012) (suggesting that ethnographers combine "emic" understandings, from the perspective of the subjects, with the researcher's own "etic" understandings rooted in their research questions and interests).

[79] *See* JULIANNE S. OKTAY, GROUNDED THEORY (2012).

Over the course of these two years of data collection, I conducted 53 semi-structured interviews. The interview participants included LPP team members, affiliates, and summer research fellows, as well legal and political advocates at other organizations working on existential risk. Some interviews were held in-person but most were remote (via video-conferencing calls) with participants who were physically located in all continents of the globe except Antarctica.[80] The interview protocol began by asking participants for a biographical account of how they became interested in the topic of existential risk. This was followed by questions about how they perceive existential risks, the community working in this area, and the organizational cultures they have encountered, including considerations of epistemics, dissent, emotions, and identity (the cultural traits discussed infra Part V). The interviews then transitioned to a discussion of legal strategy. Participants who are lawyers were also asked about how their professional identities comport with their identities as activists for the mitigation of existential risk. In addition to these formal interviews, I engaged in hundreds of hours of informal conversations with members of this community, which are not quoted or paraphrased here but nevertheless inform the empirical analysis.

With qualitative methods, the researcher serves as the research instrument in the field. In contrast to, for example, collecting public data or distributing a survey, the qualitative researcher directly asks participants their questions and observes participants in their settings, filtering the entire data collection process through the researcher's own perceptions and biases. This fact, along with a long line of critical literature about the power dynamics of representing research subjects, particularly considering the historical association between ethnography and colonialism, has led to a call for qualitative scholarship to be accompanied by a reflexive account of the researcher's interests, goals, and positionality. In this spirit, I will briefly provide relevant observations here—more auto-ethnographic notes can be found throughout the presentation of findings infra.

Over the course of the study, I transitioned from an outside observer to a more active participant in strategic discussions in this field. Following the conclusion of the ethnography, I have continued to volunteer with LPP, contributing to a range of the organization's strategic discussions. I generally agree with the notion that existential risk matters and that this community is taking useful steps toward an effective response, although I make recommendations for diversifying and scaling up this movement in Part VI. Becoming a "member" is not uncommon in ethnography, nor is it necessarily discouraged, although it can predispose the researcher to portray the community under study in a more favorable light. [81] This tendency is somewhat diminished by the existential advocates' strong cultural commitment to dissent and criticism (see infra Part V). In the LPP meeting in which I formally requested permission to conduct ethnography, I expressed my intention to contribute to socio-legal theory and to provide feedback to the organization and the larger movement around existential risk. I

---

[80] This statement assumes the convention of categorizing seven continents as North America, South America, Africa, Australia, Asia, Europe, and Antarctica.

[81] *See* Hsun-Yu Sharon Chuang, *Complete-Member Ethnography: Epistemological Intimacy, Complete Membership, and Potentials in Critical Communication Research*, 14 INT'L J. QUALITATIVE METHODS, Nov. 18, 2015, at 1, 2 (discussing the empirical value of becoming a "member" in ethnographic studies); Cf., Margaret D. LeCompte & Judith Preissle Goetz, *Problems of Reliability and Validity in Ethnographic Research*, 52 REV. EDUC. RSCH. 31, 46 (1982) (detailing the empirical and ethical risks when researchers develop strong social relationships with participants).

explained that the project would draw comparisons to how law has been deployed in other movements for large-scale change, including movements for civil rights and animal protection that I have studied in other research. LPP members embraced this project from the start and, to their credit, encouraged me to provide candid and critical feedback. I presented preliminary findings to the LPP team several times and invited participants to comment on an earlier draft of this article. This collaborative approach, in combination with assurances of confidentiality,[82] helps to promote trust and transparency in the interaction between researcher and participants. My relationship with this community has been consistently warm and respectful. This sense of rapport likely colors some of my interpretations, although my ultimate sense of responsibility as an empirical researcher is to provide an accurate description of this community, including its shortcomings and points of tension.

III. Theory of Efficacy: The Priorities Methodology

This Part begins the presentation of this article's empirical findings with a detailed summary of how these advocates approach the question of efficacy. All movements for large-scale change face the question of how to achieve lasting de facto impact. These advocates have a distinct approach to this question, which I label their approach the "priorities methodology," in keeping with the terminology of the field (e.g. the Legal Priorities Project). Their methodology has two steps. Section A explores how these advocates select goals by undertaking a global search for the cause areas where they predict that they can do the "most good" in the world. This search is continual and sometimes leads these advocates to entirely shift their focus to prioritize different cause areas and goals—including continual questioning of whether existential risk should remain a priority under this methodology. Section B explores how these advocates, upon selecting a particular objective, then make decisions about strategies and tactics. They use a "reverse engineering" framework to keep a focus on maximizing their overall, counterfactual impact toward the chosen end goal. This approach requires keeping an open mind about what strategic tools would be most useful, even if these tools sometimes have little or nothing to do with the advocates' expertise in law and policy. This zealous pursuit of impacting the most people (and other sentient beings) to the greatest extent is unique within the history of social-change lawyering. But it also raises points of tension when applied in practice. These tensions are noted in this Part and explored further in the Part V analysis of the cultural norms underlying the priorities methodology.

A. Selecting Goals

How do social-change lawyers decide to focus on a particular issue or cause? The answer from socio-legal scholarship is relatively intuitive. Lawyers, like other activists, are attracted to causes because of some combination of identity, values, ideology, and the availability of resources and opportunities.[83] Often grassroots communities already have expressed a demand for remedy or reform, which the lawyers then work to translate into legal interventions. Lawyers also pursue

[82] Identifying information is used in this article only where specific permission was granted by the identified participant.

[83] *See generally,* STUART A. SCHEINGOLD & AUSTIN SARAT, SOMETHING TO BELIEVE IN: POLITICS, PROFESSIONALISM, AND CAUSE LAWYERING (2004).

their own political values and ideologies, sometimes taking less direction from clients and constituencies. These traditional approaches can produce powerful results, as reflected in a long line of successful campaigns for civil rights and other causes. Yet, they can also lead to well-intentioned but ineffective efforts, including where lawyers fail to consider alternative cause areas where they could more effectively advance the same values that motivate them (e.g. justice, equality, democracy, well-being, and other conceptions of "the good"). This concern has led scholars of movement lawyering to call on public-interest lawyer to develop more formal processes for prioritizing cause areas and objectives.[84]

The priorities methodology developed by participants in this study provides one answer to this call. Rather than beginning with a favored cause, these advocates take the unusual starting place of "cause neutrality," meaning that they begin with an openness to working on any conceivable issue—anything in the world. This is a novel starting place for a group of public-interest lawyers. They then undergo a systematic process for selecting the cause area where they predict that they can have the greatest moral impact. As discussed below, this methodology is drawn from "Effective Altruism," a theoretical framework most often applied in philanthropy, which seeks to maximize "doing good" by combining well-meaning intentions with an evidence-based approach to effectiveness.[85]

The priorities methodology, as applied by participants in this study, ,begins with deliberations over first-principles of morality. Although this was often framed in utilitarian terms as an effort to maximize well-being, most participants described a great degree of normative uncertainty and a reluctance to fully embrace utilitarianism—with some describing themselves as "2/3 utilitarian."[86] Their next step is to determine which cause areas are amenable to the greatest impact toward chosen moral goals. The core criteria of this methodology are importance, neglect, and tractability (the "INT" analysis).[87] In this framework, a cause is prioritized not only because it affects many people and to a great degree (importance), but also because it is feasible to reduce this risk without imposing morally offsetting costs (tractability), and because the issue is not

---

[84] Cummings and Rhode, *Public Interest Litigation: Insights from Theory and Practice* (observing that public-interest law generally lacks formal processes for "identifying objectives and establishing priorities among them" and that "few organizations operate with explicitly articulated theories of change or specific measures of performance").
[85] *See generally* MACASKILL, DOING GOOD BETTER at 14; Benjamin Todd, *Can One Person Change the World? What the Evidence Says*, 80,000 HOURS, https://80000hours.org/career-guide/can-one-person-make-a-difference (Apr. 2017) ("People often wonder how they can 'make a difference.' . . . [but] the key question is, 'how can I make the most difference?'"). Effective Altruism has grown to include a number of non-profit organizations focused on research (e.g. the Global Priorities Institute; Rethink Priorities), career advising (e.g., 80,000 Hours, Training for Good), assessing effective philanthropy (e.g. GiveWell), grantmaking (e.g. Open Philanthropy, Longview Philanthropy, BERI).
[86] *See, e.g.* WINTER et al., *supra* note (citing grounds for protecting future generations primarily in utilitarianism but also in deontology, virtue ethics, and other traditions of moral and political philosophy); Transcript of *Tyler Cowen on Stubborn Attachments, Prosperity, and the Good Society*, ECONTALK (Aug. 7, 2017), https://www.econtalk.org/tyler-cowen-on-stubborn-attachments-prosperity-and-the-good-society/#audio-highlights (Tyler Cowan interview where the idea of being 2/3 utilitarian is attributed to Cowan although he says it was "tongue and cheek" https://notes.pjh.is/Two-thirds+utilitarianism).
[87] *See* MACASKILL DOING GOOD BETTER, *supra* note (outlining the "INT" analysis and proposing five key questions to assess these criteria: "How many people benefit, and by how much? Is this the most effective thing you can do? Is this area neglected? What would have happened otherwise? What are the chances of success, and how good would success be?"). AllAmericanBreakfast, *Summary Review of ITN Critiques*, EFFECTIVE ALTRUISM F. (Oct. 9, 2019), https://forum.effectivealtruism.org/posts/MtCAsPMftvJqRBYzr/wip-summary-review-of-itn-critiques.

already receiving adequate attention such that any interventions would be subject to diminishing returns (neglect).[88] For example, in recent discussions that I observed within the impact litigation team at LPP, the threat of engineered pandemics is favored by this methodology because such pandemics could produce widespread illness and death affecting current and future generations (importance), there is promising evidence that litigation could be useful in addressing issues such as unsafe laboratory practices (tractability), and the governmental budgets and overall public attention to the issue are severely lacking (neglect). For another example, this same impact litigation team has tended to view existential threats from AI as highly important and neglected but less tractable, because we do not yet have a clear enough sense of how to create safe AI, nor how litigation can contribute to this cause. This analysis could change if the AI safety research community provides more specific and high-confidence recommendations for law and policy.

Existential risk has recently emerged as a core focus of the Effective Altruist community—that is, the community applying the priorities methodology, which includes most of the advocates in this study. The prioritization of existential risk owes to the importance of the issue for a great number of current lives and a far greater number of future lives, in addition to evidence of severe neglect and at least some degree of tractability.[89] The inclusion of future generations in this analysis follows from a commitment to "moral circle expansion" and an effort to treat all sentient beings equally—whether or not these beings are especially near or similar to ourselves. In the early years of Effective Altruism (the 2000s and early 2010s), the community was focused on expanding the moral circle across spatial boundaries, leading to a focus on effective global poverty and health interventions. Empirical evidence cited in this field suggests that it currently costs roughly two to five thousand dollars, on average, to "save a life," which can be defined in different ways but in practical terms can be seen in the example of stopping a child from dying of a preventable illness where the child would then lead a full healthy lifespan.[90] The community has also applied moral circle expansion across species boundaries in recognition that non-human animals, particularly in agricultural contexts, are suffering on a large scale. And in just the past few years, this community has taken moral circle expansion across temporal boundaries, with a focus on future generations. Although it is difficult to know how our actions affect the future, existential risk is arguably less susceptible to this concern in one key respect—foreclosing the human future entirely or locking-in dystopic trajectories are impacts that persist well into the future. Although existential risk is generally favored among Effective Altruists at the moment, the priorities methodology can, by design, yield different answers. One of the foundations of this framework is a willingness to update and change directions upon new information.[91] As one

---

[88] *See* MACASKILL, DOING GOOD BETTER, *supra* note (explaining the neglect criterion by reference to the diminishing returns of investing in causes that are already receiving a great deal of attention, the potential for counterfactual impact when otherwise few resources would be spent on an issue, and the observation that the most impactful cause areas are often overlooked).

[89] Benjamin Todd, *The Case for Reducing Existential Risk*, 80,000 HOURS (Oct. 2017), https://80000hours.org/articles/existential-risks.

[90] *Why is it so Expensive to Save Lives?*, GIVEWELL (Dec. 2021), https://www.givewell.org/cost-to-save-a-life (offering the example of health interventions in Guinea, which can be estimated to save one life per $4,500 donated).

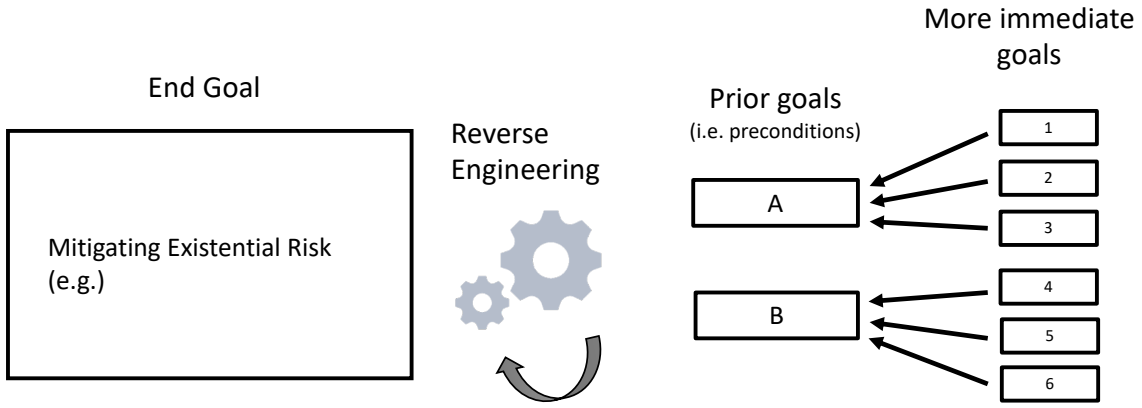
[91] WINTER et al., *supra* note ("[W]e offer a rigorous yet flexible, and potentially ever-evolving methodological framework for deciding which problems to work on and how to tackle them.").

participant explained, "If you were to prove that one of the core [Effective Altruist] cause areas actually wasn't an issue, like that AI safety was not a big risk, you would be celebrated."[92]

B. Selecting Strategies

Having selected a cause area where advocates predict that they can have the greatest impact (e.g. existential risk or particular sources of existential risk such as biosecurity relating to engineered pandemics), their next step is to select strategies in an effort to optimally advance the prioritized cause. This step is pursued through a process of reverse engineering from end goals, as visually represented in Figure 2.

Figure 2. The Prioritization Framework for Selecting Strategies



As observed in this study, these advocates with a process that I label "end goal primacy." This involves working backward from their cause (the "end goal") to more specific and immediate goals that are preconditions to achieving the end goal.[93] This approach is derived from "Theory of Change," a framework common in philanthropy and impact investing, which recommends visually mapping out the reverse engineering process.[94] In an internal document, LPP has created such a map with an end goal to advance a state of the world in which "humanity's long-term potential is safeguarded." The report notes that this goal is "far too vague to guide our decision-making" but that it helps the organization "link every subsequent step" to their end goal.[95] To

---

[92] MACASKILL, *supra* note ("[W]e genuinely just want to do what's best for the world, so if we're wrong about anything — even if it's the thing we've been dedicating our lives to — we should want to know.").

[93] LPP's website describes their approach as "mission oriented," whereby they "work backwards from these long-term goals to prioritize the projects that we think are most promising and impactful." https://www.legalpriorities.org/opportunities.html (last visited January 11, 2022).

[94] *See* Edward T. Jackson, *Interrogating the Theory of Change: Evaluating Impact Investing Where it Matters Most*, 3 J. SUSTAINABLE FIN. & INV. 95, 100 (2012) (recommending visual mapping to identify "underlying logic, assumptions, influences, causal linkages, and expected outcomes of a development program or project").

[95] Other organizations working on the long-term future of humanity have also applied elements of Theory of Change. For example, the Simon Institute for Longterm Governance published a diagram that lists an end goal of "long-term human flourishing" with preconditions to improve "long-term institutional fit" through improving

take the example of just one causal chain through this document, they describe a condition where research on existential risk is "known and valued by policy-makers." This is preceded by building relationships with policy makers and encouraging LPP affiliates to pursue policy careers, which is preceded by building relationships with organizations on the frontlines of policy outreach and advocacy. The key point here is that more specific and immediate goals are defined after higher goals. If new information alters higher goals, this model requires a willingness to reconsider an organization's immediate and day-to-day strategic priorities.

When designing strategies, the existential advocates draw insights from the "integrated advocacy" literature on law and social change.[96] Scholars in this field recommend that advocates blur the distinction between law and policy, coordinating strategies and frames across the traditional law/policy divide.[97] For participants in this study, this integration is consistent with the Effective Altruist notion of a "portfolio approach" and an "alliance mentality," emphasizing the overall impact of collective efforts toward a particular end goal, rather than the isolated impact of the actions taken by any particular individual or organization.[98] An internal LPP report on impact litigation directly references the integrated advocacy literature and emphasizes that litigation efforts should be paired with complementary actions in the domains of legislation, policy, public opinion, academic research, and education.

The scholars of integrated advocacy view the blurring of law and policy as a corrective to the long-standing concern that lawyers in social movements tend to over-emphasize the value of law and courts, falling under a "myth of rights" while discouraging legislative outreach, grassroots organizing, and other forms of activism that might prove more effective.[99] But the myth of rights does not seem to hold much appeal among the existential advocates. Some participants in this study noted that, until very recently, lawyers in Effective Altruism were more drawn to the opposite myth: that, for Effective Altruist causes, as one participant put it, "litigation is useless, law is useless."[100] Until recently, law students who were committed to Effective Altruism were primarily advised to take high-paying positions so that they could "earn to give" to effective charities rather than seeking out impactful legal work relating to prioritized cause areas. This trend has changed over the past few years. The Effective Altruism community now appears to see much more value in legal and political efforts. Some hesitations about the role of law remain,

---

decision-making process and integrating longtermist concerns in "dominant societal narratives," "institutions," and "policy agendas." https://www.simoninstitute.ch/blog/post/our-theory-of-change/ (last visited August 12, 2022).

[96] *See* Cummings, *supra* note (observing that the "new convention" among public interest lawyers is to engage in "multidimensional" and "integrated" advocacy where lawyers are "strategically sophisticated" and work with a wide range of allies to "advance political goals in multiple venues through coordinated tactics in the face of persistent opposition").

[97] See Cummings, *supra* note (reviewing the interdisciplinary literature on the "interaction between law and politics" and recommending that law be "coordinated with politics through an integrated strategy that maximizes the potential for sustainable social change.").

[98] *See* MacAskill, *supra* note ("The fact that we each act as part of a wider community warrants a 'portfolio approach' to doing good—taking the perspective of how the community as a whole can maximize its impact.").

[99] *See* SCHEINGOLD AND SARAT, *supra* (noting that social-change lawyers tend to be "attracted to courts as fly paper," mesmerized by a "myth of rights").

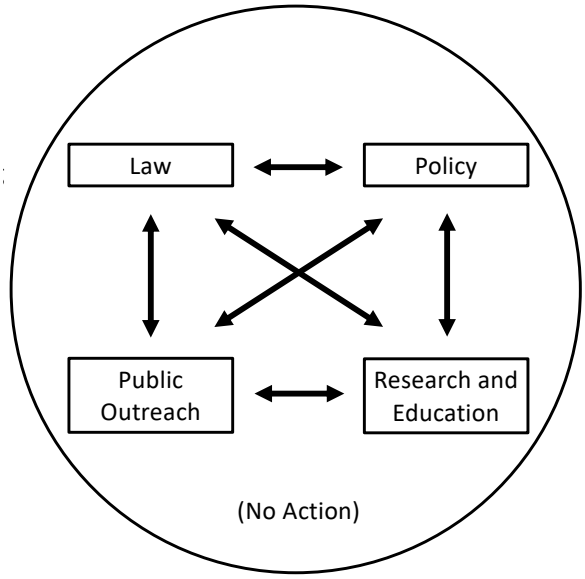
[100] Participants working in legislative outreach made very similar observations about Effective Altruists' historical lack of interest in the policy domain. For example, one such participant observed that when they attended Effective Altruism conferences just a few years ago they would rarely find anyone with an "interest in living in DC or working with the US government."

although the legal efforts in this space are receiving a great deal of funding and support. The persistence of these hesitations may help this community avoid falling into a myth of rights. For example, these advocates are considering efforts to create justiciable rights for future generations in domestic constitutions, common law doctrines, international law (e.g., extensions of human rights across time), and inter-governmental agreements. But participants were consistently mindful that simply establishing such rights would not necessarily mean they would be enforced in transformative or otherwise meaningful ways.

A central commitment of the priorities methodology is to undertake counterfactual analysis, assessing the marginal expected value of any action under consideration.[101] This involves weighing how one option compares to alternatives, including the possibility of doing nothing.[102] This approach draws from the Effective Altruist norm of asking "what would have happened otherwise." In my observations at LPP, this counterfactual approach was consistently evident in debates over planned legal interventions. As described in an LPP report, any strategic action should be compared to "feasible alternatives and counterfactuals," including consideration of "the chances the legislature or executive will take up this issue at some point anyway" or whether these governmental entities will act "with enough haste."[103] For example, when considering whether to provide formal comments on the UN Declaration on Future Generations, LPP engaged in a lengthy decision-making process about whether their input would be counterfactually impactful when considering what would happen otherwise and how the same "people hours" could be spent on other projects. LPP discussed this issue during a four-day in-person "theory of change" retreat that I attended. This emphasis on counterfactuals is represented by the term "no action" in Figure 3, which summarizes the understanding of integrated advocacy among the participants in this study.

[101] *See* WILLIAM MACASKILL, DOING GOOD BETTER 57 (2015) (describing "marginal utility" as a "fundamental piece of scientific reasoning" that motivates the methodology of Effective Altruism).
[102] As one participant explained, legal actions should be evaluated according to the extent to which they have an "independent or a counterfactual and positive impact."
[103] LPP internal litigation report, on file with author.

Figure 3. Integrated Advocacy



When seeking out strategies that maximize their impact, these advocates give considerable weight to how their efforts will deal with cognitive biases. This is evident in discussions of specific strategies as well as high-level discussions about whether to focus more attention on law (e.g., litigation and judicial outreach) or policy (e.g. legislative and regulatory advocacy). A participant working in the policy space emphasized that their work is often stymied by the range of cognitive biases that inhibit recognition of existential risk, noting that "99.5% of policy makers don't actually realistically think in the next 100 years human beings could possibly go extinct…." Another participant working in this space emphasized that elected officials are focused on the near-term demands of their constituents and, more generally, have "no time" because they have people "hounding them…every second of the day" and an incessant "huge box of materials to go through." Thus, when presenting existential risk to policymakers, one participant reported commonly receiving the response: "How can I think about this when we've got so many crocodiles closer to the boat?"[104] Participants working on the legal dimension of these issues cited further law-specific biases, including the notion that law tends to have a conservative, backward-looking precedential orientation that might be unreceptive to the "new legal techniques" that would help to address existential risks to future generations.[105] Moreover, participants worried that the judiciary might struggle with the probabilistic nature of existential risk given judges' lack of "formal training in quantitative subjects."[106] But participants also saw promise in the judiciary to care for future generations as a disenfranchised group. As one

[104] As one participant working in policy outreach explained: "Most people in political realm are working on 24-hour review cycle. There's a lot to be done in the short-term, so its hard to get anything done in the long-term. Long-term is a year out from now. That's how chronologically structured political folks think about things. But with respect to 100 years that would be a really hard case to make."
[105] *See* Albiston, *supra*.
[106] WINTER et al., *supra* note (noting that law tends to rely on non-quantitative reasoning, as evident in standards framed as "beyond a reasonable doubt," "probable cause," and "balancing tests.")

participant noted, courts uphold the "liberal values" of democracies and the protection of the "political minority." Several participants expressed optimism that judges, owing to their general reliance on "abstract values," might be receptive to expanding human rights "independently of time." It is also important to note that existential advocates are taking some actions that directly seek to reduce cognitive biases, including holding educational workshops for policymakers (and considering similar workshops for judges) on existential threats and the distinct decision-making challenges around low-probability/high-impact risks.

So far, this section has described the ex ante phase of the priorities methodology—predicting how planned actions might optimally advance an end goal. But this effort to maximize impact also demands continual empirical assessment of actions as they are taken. This involves a commitment to empiricism, Bayesian updating, and developing metrics and experimental methods when possible.[107] But these advocates also draw on some more familiar considerations from socio-legal literature, which emphasizes different ways of conceiving of social-change impact that follows from legal activism. As applied by participants in this study, this includes questions of whether to pursue broad and holistic approaches to existential risks (e.g. establishing rights and legal standing for future generations, criminalizing activities that generate existential threats, influencing regulatory policy that affects multiple branches of government) or more narrow and cause-specific legal challenges (e.g., enforcing regulations on a laboratory developing a dangerous technology in an unsafe manner). They also follow the socio-legal distinction between direct effects (e.g,, how a court order is enforced) and indirect effects (e.g., how a court order reshapes political discourse). As described in LPP internal documents, some legal actions may be especially valuable for their indirect effects, for example where a litigation victory or loss serves an educational function (e.g., to reveal the urgency of a problem), a motivational function (e.g. what socio-legal scholars call "internal effects" where a legal victory helps stimulate growth in the movement), a symbolic function (e.g. legitimating ideas in the movement, and gaining control over issue definition), and a resource mobilization function (e.g., attracting funding and increasing bargaining power). These indirect effects may be particularly important in the context of international soft law, which participants consider a key avenue for addressing the global dimension of existential risks.[108] As one participant noted, even if soft law does not have the teeth of strong enforcement mechanisms, it can help "inform the common language that people use…how they perceive the future…and whether or not they consider extreme risks."

[107] An internal LPP report makes the case that "specific measures of performance" are necessary to assess "progress…made towards our goals in an observable, measurable way" and to "identify where theory and practice fall apart and thus where we should rework our [Theory of Change]." Experimental projects in this field include a forthcoming study of how judges may interpret legal arguments relating to existential risk cause areas, manipulating a number of factors including source and content of law, level of expert disagreement, how broad the rule is, and the level of abstraction at which a choice is presented.

[108] Participant acknowledged the limitations of international soft law. For example, one participated cited the failures of international health regulations, which they described as "the most advanced form of regulation in the international system" and as "theoretically binding," but noted that these regulations were "not respected during Covid."

IV. Theory of Accountability: Representing Future and Current Generations

The previous Part described the priorities methodology as a formula for enhancing efficacy. But if this community were entirely focused on maximizing impact, they might tend to overlook important ethical considerations relating to their inclusiveness and accountability. Moreover, these ethical dimensions have been a salient concern in the literature on social-change lawyering. Even the most celebrated civil rights advocates have been charged, in some scholarly accounts, with failing to hear or heed the voices of the populations most affected by an issue.[109] Recent scholarship acknowledges this "elite critique" but tends to offer an increasingly collaborative portrait of activist lawyers, in many contexts, working alongside movements and seeking to secure the remedies desired by key constituencies.[110] These debates about the accountability of lawyers find new expression among the existential advocates, who conceive of their primary constituency, essentially their clients, as the multitudes of people who could exist in the future.[111] Some participants described themselves as part of the tradition of civil rights, protecting future generations in an effort to, as one participant put it, "give voice to people who have been underrepresented."[112] But how can these lawyers be accountable to a population that does not (yet, and might not ever) exist, and so cannot speak? Moreover, accountability in this context is complicated by the question of how to include the various perspectives of currently living people—a population that is also affected by existential risk and the costs of interventions. This raises the central puzzle explored in this Part: How do these advocates conceive of their accountability to current people (who are capable of speaking) and future people (who are potentially the greater affected population but incapable of speaking)?[113]

Accountability to clients and constituencies is a well-developed topic in the literature on the professional responsibility of lawyers. First-order representation issues, which relate to duties to clients, raise an inherent tension between the lawyers' justified paternalism, owing to their expertise, and clients' justified autonomy and control over objectives.[114] In the context of social-change lawyering, these tensions can be heightened where attorneys prioritize the broader impact

---

[109] *See* Derrick A. Bell, Jr.*, Serving Two Masters: Integration Ideals and Client Interests in School Desegregation Litigation,* 85 YALE L.J. 470, 489–91 (1976) (noting that the lawyers of the Civil Rights Movement chose to prioritize racial integration in schools while Black Southerners expressed a clear preference to focus on educational quality and other approaches to addressing racial subordination in schools).

[110] See CUMMINGS, *supra* note.

[111] It is worth noting that many participants in this study would frame the issue as a concern not only for future humans but also for future non-human animals, possible post-humans, and possible sentient AI.

[112] *See also*, ORD, *supra* note (calling for efforts to bring "the representation of future generations into national and international democratic institutions.").

[113] Note that the notion that future persons are the greater affected population implies either existential catastrophes that leave future persons to live under dystopic conditions or, under a human extinction scenario, a belief that the existence of potential future persons matters (i.e. that human extinction would cause harm beyond ending the lives of the people who would experience the extinction event). See discussion supra Part I.

[114] *See* Judith Maute, *Allocation of Decisionmaking Authority Under the Model Rules of Professional Conduct*, 17 U.C. DAVIS L. REV. 1049, 1081 (1984). (paternalism vs. client-service vision of the profession); Marcy Strauss, *Toward a Revised Model of Attorney-Client Relationship: The Argument for Autonomy*, 65 N.C. L. REV. 315 (1987); Mark Spiegel, *Lawyering and Client Decisionmaking: Informed Consent and the Legal Profession*, 128 U. PA. L. REV. 41 (1979).

of a case over securing remedies for the particular client.[115] Participants in this study anticipated first-order issues, noting that their litigation plans may tend to focus on relatively remote (future) impacts of a case, which may be orthogonal or opposed to the interests of an individual client.[116]

But participants were more troubled by representational issues of the second order, regarding duties to constituencies, causes, and a sense of "public accountability." These second-order duties are largely absent from the rules of professional conduct, except in aspirational prefatory language where the U.S. Model Rules describes the lawyer as a "public citizen having special responsibility for the quality of justice," who should "seek improvement of the law" and engage in "civic influence."[117] This concern for the public good is especially central to the practice and theory of "cause lawyering," which refers to lawyers who work on behalf of causes and social movements.[118] The image of the lawyer as an amoral hired-gun for clients is replaced in this tradition by the lawyer's own "political or moral commitment."[119] In a wide range of movement lawyering contexts, this approach has raised difficult questions about how, and to what extent, key constituencies should be included in strategic decision-making processes.[120]

The participants in this study seem to fit squarely within the cause-lawyering tradition, drawing from their moral imperative to "do the most good." LPP members regularly deliberated over second-order representational issues. In one meeting, I suggested potential labels for LPP's theory of accountability, such as "first-principle accountability." This framing stresses moral commitment, although participants noted that the term lacks recognition of their methodology for maximizing impact. One participant suggested "internal accountability," as opposed to "external accountability," where "you have direct feedback from living constituencies." This participant explained that focusing on future generations, who cannot provide input, means that existential advocates movement must find accountability through the integrity of their own methodologies and practices. Regarding the commitment to the interests of future generations, I proposed the term, "astronomical value accountability," in reference to the arguably great value associated with future persons—with a future potential of perhaps 10^16 "human lives of normal duration," in addition to the likelihood that technological developments could greatly increase this number if we do not destroy ourselves along the way.[121] Participants were unpersuaded by this label, in

---

[115] *See* Susan Carle & Scott L. Cummings, *A Reflection on the Ethics of Movement Lawyering*, 31 Geo. J. Legal Ethics 447, 460–61 (2018) (discussing conflicts that arise between the client's interests in a speedy and beneficial remedy and the lawyer's interests in advancing a larger cause or reform).

[116] An internal LPP report notes that a "long-term cause lawyer" may face conflicts of interest where they are forced to "choose between the cause and the client." The report further notes a desire to "avoid the type of victimization" that has occurred in some civil rights litigation where plaintiffs' interests have been overridden or "misrepresented."

[117] MODEL RULES OF PROF'L CONDUCT Preamble.

[118] *See* Stephen L. Pepper, *The Lawyer's Amoral Ethical Role: A Defense, A Problem, and Some Possibilities*, AM. BAR FOUND. RES. J. 613 (1986) (defending the "standard conception" of the lawyer role, which demands zealous partisanship on behalf of the client's interests while holding in abeyance the lawyer's own moral assessments of client objectives).

[119] *Id.* STUART A. SCHEINGOLD & AUSTIN SARAT, SOMETHING TO BELIEVE IN: POLITICS, PROFESSIONALISM AND CAUSE LAWYERING 4 (2004).

[120] Cummings forthcoming, *supra* note.

[121] Nick Bostrom, *Existential Risk Prevention as Global Priority*, 4 GLOB POLICY. 15 (2013); BOSTROM AND CIRKOVIC, *supra* note (discussing scientific predictions that complex life on earth may last for .9 to 1.5 billion years, although it is possible that space travel could enable us to survive until the last stars burn out, around 100 trillion years for now, or even until black holes disappear, which might be 10 ^100 years from now, or even longer).

part because it seems to imply "strong longtermism," the view that our greatest priority today should be impacting the far future.[122] Most participants expressed uncertainty and skepticism about this brand of longtermism. A less "strong" version of "future generation accountability" might be a more accurate term, reflecting the general sense of compassion and concern that participants expressed for people who will one day exist and have meaningful lives if we can avoid an existential event. But this term fails to capture the concern that participants consistently expressed for current living people who are affected by existential risk—and the overlapping issue of sub-existential catastrophic risk.

Within this discussion, some participants were skeptical of any suggestion that they "represent" future generations, because this would imply a fiduciary relationship that seems impossible given the inability to receive communications from the parties being represented.[123] Advocates in this field reference an "epistemic challenge" when seeking to know what future generations might want and need. All social-change lawyers face some degree of epistemic challenge when seeking to understand what an affected constituency might desire, especially when representing loosely affiliated groups with complex power dynamics.[124] This epistemic problem is arguably even more difficult when working to protect the long-term future. History reveals vast changes in human values over centuries and millennia, which suggests that any values that we might lock in today, no matter how beneficial they seem to us, might be deemed undesirable by future generations. Participants responded to this challenge by suggesting that future generations would want us to protect their "basic needs," as defined in the lower third of Maslow's hierarchy, such as the necessities of survival and a baseline of well-being.[125] Another common response among these advocates was to emphasize the principle of optionality, which recommends avoiding locked-in and irreversible effects that reduce the autonomy of future generations. The mitigation of existential risk would seem to fit both of these responses—by avoiding extinction and permanent dystopia, these advocates hope to secure basic needs and preserve optionality.

A key point of discussion among these advocates is the question of how much accountability is owed to people alive today. Some participants de-emphasized this notion of current-person accountability. As one participant explained, to follow the prevailing norms and values of "present humans" may be "almost irresponsible" because humans generally are subject to cognitive biases against recognizing existential risks and the moral interests of such a distant population as future generations. Another participant argued that conceiving of accountability primarily to future rather than current persons is more inclusive than prevailing theories of democracy, which can be critiqued for "only taking into consideration the interests and preferences of people currently alive."

---

[122] Greaves and Macaskill, *supra* note. Cite *The Case for Strong Longtermism* (defining strong longtermism as "the view that impact on the far future is the most important feature of our actions today").
[123] *See* MACASKILL, *supra* note ("Though we cannot give genuine political power to future people, we can at least give consideration to them. Abandoning the tyranny of the present over the future, we can act as trustees.").
[124] ANN SOUTHWORTH, $PEECH: THE CAMPAIGN TO UNLEASH BIG MONEY IN AMERICAN POLITICS, forthcoming (noting that the rules of professional conduct are silent on how movement lawyers should represent "groups with poorly defined decision-making processes, or loose coalitions in which power among the different groups within the coalition is unequal…[where] it can be difficult to decide who speaks for the group or coalition, and it can be hard to reconcile the competing claims").
[125] *See* Saul McLeod, *Maslow's Hierarchy of Needs*, 1 SIMPLY PSYCH. 1 (2007).

But most participants, including those who greatly value the interests of future generations, also emphasized the importance of being accountable to current persons. These advocates are working to mitigate threats that could cause a great deal of suffering and loss of life to people alive today.[126] Participants regularly discussed how these harms, if realized, would likely be distributed unequally, along the familiar axes of global inequality.[127] Moreover, these advocates have proposed new governing structures that would bring a broader community of current persons into the efforts to protect future persons, e.g. the proposed creation of legislative panels of randomly selected citizens with a mandate to protect future generations.[128]

Some participants worried that if they were only accountable to future generations, the distant and abstract nature of this population might tend to undermine the notion of accountability altogether, leading to a self-interested and biased application of the priorities methodology. Participants sometimes referenced current-person accountability as an analogy to help internalize their sense of future-person accountability, e.g. noting that the $5,000 spent on an event focused on protecting future generations could have instead been spent to save a child's life who would otherwise die from malaria. This creates an urgent sense that efforts aimed at the future should be subjected to a great deal of scrutiny because lives are at stake.

These discussions of current-person accountability often centered on the question of who seems to have a seat at the table and who seems to be excluded in the community of existential advocates. This was usually framed as an issue of diversity. Although this movement started with an Oxford-based conversation among mostly white men, the community has diversified over time.[129] Relative to many other organizations in this field, LPP appears to be quite diverse in terms of international representation. In addition to their wide-ranging geographic backgrounds, LPP adopts a cultural value to "think globally,"[130] which is reflected in their strategic discussions (e.g. planning legal interventions according to their cross-jurisdictional impacts), as well as informal conversations (e.g., remarking on cultural differences as revealed through anecdotes from day-to-day life or different reactions to global news events). As an ethnographer and interviewer, the global nature of this movement was a near constant theme in my observations,

---

[126] *See supra* Part I.

[127] An intenral LPP report emphasized the inequalities often associated with catastrophic events, as some harms "may be readily avoided and mitigated locally by those with resources."

[128] MacAskill and John, *supra* note (proposing citizens' panels as a "novel representative, deliberative, and future-oriented body" with "an explicit mandate to represent the interests of future generations").

[129] *Compare* Neil_Dullaghan, *EA Survey 2019 Series: Community Demographics & Characteristics*, EFFECTIVE ALTRUISM F. (Dec. 5, 2021), https://forum.effectivealtruism.org/posts/wtQ3XCL35uxjXpwjE/ea-survey-2019-series-community-demographics-and (reporting a survey of the Effective Altruist community showing that respondents disproportionately identify as male (71%), white only (87%), and young (median age of 28, mean 31, and 78% younger than 35)) *with* David_Moss, *EA Survey 2020: Demographics*, *supra*, https://forum.effectivealtruism.org/posts/ThdR8FzcfA8wckTJi/ea-survey-2020-demographics (reporting a survey of the Effective Altruist community showing that respondents still disproportionately identify as male (70.5%), white only (75.9%), and young (median age of 27, mean 29, and 80% younger than 35)) *See also* Vaidehi Agarwalla, *2019 Ethnic Diversity Community Survey*, EFFECTIVE ALTRUISM F. (May 11, 2020), https://forum.effectivealtruism.org/posts/2T3cGecjHfbEPXeEc/2019-ethnic-diversity-community-survey; Anonymous, *Anonymous Contributors Answer: How Should the Effective Altruism Community Things About Diversity?*, 80,000 HOURS (Apr. 27, 2020), https://80000hours.org/2020/04/anonymous-answers-diversity (discussing measures of diversity in the Effective Altruism community).

[130] *Open Positions*, LEGAL PRIORITIES PROJECT, https://www.legalpriorities.org/open-positions.html (last visited Jan. 18, 2023) (noting that their "staff is spread around the world.").

whether traveling in different countries or engaging in remote meetings held at odd hours to accommodate different time zones. This movement's ability to operate on a global level is facilitated by the widespread rise of remote work, and the related technological tools that have developed, accelerated by the COVID-19 pandemic. While this makes for a diverse community in some respects, participants still emphasized major representational deficits, e.g. the community continues to overrepresent the Global North and white men.

Diversity is valued in this community primarily for the sake of inclusivity, i.e. opening the discussion around existential risk to a wider array of voices. But it is also possible that by expanding current-person representation, this movement may grow more effective. A more heterogeneous community may yield information that would otherwise be overlooked, and this may serve the movement's goal of maximizing impact.[131] Moreover, participants noted that cultivating a broader range of voices may be essential to persuading lawmakers and policymakers to support important actions relevant to existential risk. To take an example from the international policy efforts, participants observed that some Global South diplomats appear somewhat resistant to proposals relating to existential risk, both because such proposals seem to divert attention away from issues currently affecting their constituencies and due to general distrust of the Global North countries where theories of existential risk have originated. One response to this issue may be found in LPP's support for the development of new existential risk initiatives and Effective Altruism fellowships in the Global South.[132] The executive director of LPP is a law professor at Instituto Tecnológico Autónomo de México, where he has developed existential risk programming. Another LPP member from Kenya recently led an 11-week fellowship with a group of around 60 law students in Nairobi. Although this participant reported that some fellows initially found longtermism "almost ridiculous…given the problems we have today of poverty and hunger…and corruption," his survey at the end of the term showed that the fellows' "minds were changed drastically" and they rated concerns for "far future people" as the cause area deserving of the greatest concern. This participant observed that several fellows from this program have continued to engage with the topic of existential risk. If this movement can shift to Global South leadership, participants noted that this may help to create a more effective and persuasive demand for legal and political action.

---

[131] Holden Karnofsky, *Worldview Diversification*, OPEN PHILANTHROPY (Dec. 16, 2016), https://www.openphilanthropy.org/research/worldview-diversification; Luke Freeman, *"Big Tent" Effective Altruism is Very Important (Particularly Right Now)*, EFFECTIVE ALTRUISM F. (May 19, 2022), https://forum.effectivealtruism.org/posts/SjK9mzSkWQttykKu6/big-tent-effective-altruism-is-very-important-particularly (defining the "Big Tent" approach of EA community building as one that "encourages 'a broad spectrum of views among its members.'").

[132] The first Effective Altruism conferences in Latin America and India were held in January 2023. *See https://www.centreforeffectivealtruism.org/blog/latin-america-and-india (last visited* Jan 27, 2023).

V. The Culture of Existential Advocacy

Having discussed the theories of efficacy (maximizing moral impact through a priorities methodology) and accountability (with a focus on representing the interests of future generations), this Part examines how such a model is realized in the daily culture of social-change lawyering. One might assume that the priorities methodology and the commitment to future generations could serve as an ideal but would find little expression in practice. This model seems to demand a commitment to evidence-based reasoning, while setting aside other considerations, biases, and incentives. It also demands a commitment to working on behalf of a future population that is invisible and difficult to even imagine. Yet, at least in this early stage of their movement, participants seem remarkably adherent to a set of scientific truth-seeking norms that support their methodology and their focus on future generations. As categorized in the findings presented below, these norms relate to uncertainty (Section A), deliberative rationality (Section B), supportive dissent (Section C), and limiting group identity (Section D). Throughout the presentation of these findings, I note that these scientific norms conflict with some aspects of human nature and participants' identities as lawyers and as members of a social movement.

A. The Uncertainty Norm

Throughout the interviews and ethnographic observations in this study one of the most pervasive cultural themes was uncertainty. The notion of existential risk is fundamentally a matter of uncertain and quite speculative risk assessment regarding possible future events. This differentiates existential advocates from movement lawyers in many other contexts, where (i.e. in a movement to counteract discrimination against some class of current living persons) the notion that some harm is occurring is viewed as a certainty. For the existential advocates, their goal is to reduce the probability of a particular class of a particular class of events. One participant noted that a reduction from a per-century 15% chance of existential catastrophe to 10% would be a major win for this movement. Thus, success looks like something not happening. Failure could similarly look like nothing has happened in a sense, where failure means a catastrophic loss of life such that the advocates would not be there to experience the outcome. Moreover, it can be difficult to know how much risk reduction has been achieved and how much can be attributed to the work of advocates. Even more fundamentally, participants expressed a great deal of uncertainty, both normative and empirical, about how much global priority should be placed on existential risk. In my ethnographic observations, I was struck by how often participants engaged in long discussions of the question of whether or to what extent existential risk matters (i.e. how likely threat scenarios are and how much to discount harm to future generations).

Uncertainty was also a central theme when participants applied the priorities methodology, which can be viewed as an effort to estimate impact based on available evidence and best guesses. This methodology requires maintaining a sense of uncertainty so that one can continually reassess priorities rather than getting too attached to a particular conclusion about goals or strategies. The notion of "updating" views was pervasive in the vocabulary of this community, as participants often described changes to their views upon receiving new information or having a new discussion of a topic. This commitment to updating requires maintaining a sense of uncertainty. Thus, rather than viewing uncertainty as an obstacle, this

community tends to view it as something to embrace as a cultural norm.[133] As one participant put it, a core objective of this community is to "normalize uncertainty" both within existential advocacy organizations and in their outward-facing advocacy and educational efforts. Members of this community regularly remind one another to maintain a sense of uncertainty, often admonishing peers who seem to overstate the confidence in their claims. One participant explained this cultural norm by observing, "anyone who comes across as overly certain will be greeted with suspicion." This norm was often framed as a matter of "epistemic humility," which is a foundational concept in Effective Altruism. For example, posts on the Effective Altruism Forum generally begin with an "epistemic status," as recommended by the Forum guidelines, describing the degree of confidence the author has in their empirical and normative claims.[134]

In order to pursue strategic action under these conditions of uncertainty, and these norms that continually draw attention to uncertainty, the existential advocates look to expected value theory, approaches to low-probability/high-impact events in decision theory, and Bayesian reasoning.[135] But participants acknowledged that these rational foundations of their "uncertainty culture" are difficult to realize in the face of the general cognitive tendency to seek out more firm understandings. This tendency may be heightened in the context of advocacy, e.g. where one is seeking to persuade lawmakers that they should address real, even if uncertain, problems. In such contexts, participants noted that they often need to de-emphasize the language of uncertainty, effectively code switching. As one participant noted: "if you speak the language of EA epistemics within policy culture, they will not like it… you need to speak a different dialect with them." Participants also emphasized that too much emphasis on uncertainty can have some negative effects within organizations working on existential risk. Uncertainty can inhibit action, where strategic discussions digress and become unduly complicated or lead to changing directions too often. As one participant explained, "in the end, one has to commit to a certain path at least for some time… [rather than] changing your trajectory every few weeks."

B. The Deliberative Rationality Norm

The priorities methodology is rooted in the notion of combining the head with the heart—but putting a little more emphasis on the head to assure that the heart is guided toward doing the most good. This approach depends on a highly deliberative "System 2" form of cognition.[136] While strong emotions and automatic, instinctive cognition can be instructive in thinking about what matters, participants worried that drawing too heavily on "System 1" could lead to focusing on the most familiar and personally relevant cause areas while overlooking opportunities for greater impact. It is conceivable that strong emotions would drive efforts to mitigate existential

---

[133] *See, e.g.,* Keiran Harris, *Effective Altruism in a Nutshell* (Oct. 18, 2021), https://80000hours.org/2021/10/effective-altruism-in-a-nutshell/ (emphasizing the importance of uncertainty and humility within Effective Altruism, and conceding, "maybe the sceptics are right, and we're just wasting our time.").

[134] A typical example reads: "Epistemic Status: Quickly written (~4 hours), uncertain. [This topic] is not my field of expertise." These notes are typically accompanied by symbols, e.g. (+) (++) (-) (--) (?) (??), which depict degrees of uncertainty and other valences.

[135] *See* Bostrom, *supra* note (introducing a "maxipok" principle, which aims to maximize the probability of not having an existential disaster, and thus having an at least "ok" outcome); MACASKILL, *supra* note.

[136] *See generally*, DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2011) (describing "system 1" cognition as emotional and instinctive ("fast") and "system 2" cognition as more rational and deliberative ("slow")).

risk, such as fear about apocalyptic scenarios, anger toward particular entities that exacerbate risks, or hope for visions of a utopian future. But these framings were uncommon and disfavored among participants. Some even reflected on their own "missing mood," wherein their concern for the "long-term future of billions and billions of people" is not met by a commensurate emotional response. This missing mood may be a product of the cognitive biases that fundamentally limit our ability to comprehend the scale of existential risks (as discussed supra Part I). Even for the participants in this study who work with existential risk on a daily basis, it is difficult, as one participant put it, to "emotionalize uncertainty" regarding the likelihood of existential events and to generate strong feelings for "people who don't exist yet," because "you can't meet them" and 'you can't have a graphic documentary about them." One participant contrasted this missing mood with their work in animal welfare, where advocates would frequently "cry together" during meetings,[137] and in social justice contexts, where this participant remarked that they "miss the…bleeding heart of being more emotionally drawn to the cause."[138]

But the missing mood in this community should not be exaggerated, as some participants described their work in more emotional terms. For example, one participant explained that they were drawn to working on existential risk after undergoing extensive psychotherapy to get more "connected to [their] emotions" and to have more "emotional capacity for compassion with broader groups and issues," which led to thinking about how they could "care about other people in a meaningful way" and then whether there are "more and less effective ways of doing this." Two participants reported drawing a poignant source of motivation from the 2020 entreaty of a suicide note written by a Harvard Law School Effective Altruism member, who wrote, as reported by family members in the press: "Please look after each other, the animals, and the global poor for me."[139] One participant described their admiration for this late colleague's deeply felt and expansive compassion and "beautiful heart."

The understanding of emotional reasoning in this field is sometimes framed as question of rationality. For example, I observed an internal LPP discussion over whether to continue using the term "rational" in the public description of their organizational culture, which, at the time, read: "Rational: We use evidence and careful analysis to tackle the world's most pressing problems…" Some members worried that this term may imply a simplistic rational/emotional dichotomy and a judgement toward others for being "irrational,"[140] although some conceded that the term accurately describes "an interest in being good with thinking, good with statistics," and an effort to encourage "Bayesian thinking" and "ways of reducing cognitive bias." This terminological debate reflects a deep tension regarding the role of emotions in motivating this

---

[137] As this participant noted, "seeing animals suffer is emotionally extremely charged, and seeing animals happy is also emotionally charged."

[138] Another participant similarly noted, "I think it is infrequent that I'm super connected to the emotions of 'oh no, how horrible would it be if this [catastrophic event] really happened?' Compared to being involved in the social justice movement where I can see these things and experience them in my whole life…"

[139] *See* Meagan Flynn, *Rep. Raskin and his Wife on Their Late Son: 'A Radiant Light in this Broken World,'* THE WASHINGTON POST (Jan. 4, 2021), https://www.washingtonpost.com/local/md-politics/rep-jamie-raskin-and-wife-sarah-share-moving-tribute-remembering-their-son-tommy-raskin/2021/01/04/0ef01b30-4ee3-11eb-83e3-322644d82356_story.html (noting that the deceased was a committed and vocal board member of Harvard Law School Effective Altruism).

[140] *See* James M. Jasper, *Emotions and Social Movements: Twenty Years of Theory and Research*, 37 ANN. REV. SOCIOLOGY 1 (2011) (challenging the rationality/emotion dichotomy with the observation that "feeling and thinking are parallel" and are "composed of similar neurological building blocks").

community. Effective Altruism is often said to require "not going with your gut" and instead "taking a step back, figuring out your values, and determining where you can have the most impact."[141] This framework scrutinizes emotional reasoning, in recognition that often "choosing from our heart is unfair," and it is crucial to "do the math, to go find the numbers…figure out how many people a problem affects…[and] how badly it affects them…"[142] In *Doing Good Better*, MacAskill recommends "combining the heart and the head."[143] This formulation was perhaps an easy fit with MacAskill's focus, at the tim, on current-time causes of global health and animal welfare. Combining the heart and the head might be more difficult in the context of existential risk, which raises the less emotionally salient issue of protecting future generations. Some participants sought to resolve this tension by finding "outlets for the heart" in their pro bono and philanthropic contributions outside of their full-time daily efforts to reduce existential risk. One such participant explained that they felt more "emotional connection" to their donations to effective global poverty interventions, because: "my heart is there [with global poverty], and my head is with longtermism."

C. The Supportive Dissent Norm

These two cultural underpinnings of the priorities methodology already discussed—uncertainty and deliberative rationality—are further supported by norms of dissent and heterogeneous discourse. Dissent is an express pillar of the Effective Altruism framework.[144] In the effort to determine what priorities and strategies can be expected to maximize impact, dissenting views can help reveal uncertainties and provide information for deliberative processes. The dissent norm was often framed as a matter of limiting "value alignment" that would lead to excessive groupthink about priorities.[145] Although members of the Effective Altruist community often discuss the extent to which someone is "EA aligned," many participants in this study seemed uncomfortable with the homogeneity implied by this terminology—two participants noted that "alignment" has even worse connotations in languages other than English where it implies something like marching in formation.

Some of the non-profit organizations in this field provide explicit reward functions for evidence that tends to disconfirm conclusions about the organization's priorities. This includes "red team challenges," where prizes are awarded to the best criticism and counter-arguments regarding common conclusions in the Effective Altruism community.[146] In the meetings that I observed,

---

[141] MACASKILL, *supra* note at 10 (arguing that "relying on good intentions alone to inform your decisions is potentially disastrous").
[142] Cotra, *supra* note.
[143] MACASKILL, *supra* note (noting that the Effective Altruist notion of "combining the heart and the head" can mean that the heart inspires altruistic pursuits and the head informs those pursuits by turning "good intentions into astonishingly good outcomes").
[144] Cotra, *supra* note; *see also,* MacAskill, *supra* note (noting that the Effective Altruist culture of dissent and "independence of thought" is reflected in the observation that several of the "most-upvoted Forum posts" are "critical" or "critically self-reflective" in nature).
[145] *See generally,* Carla Zoe, *Objections to Value-Alignment Between Effective Altruists* (Jul. 15, 2020), https://forum.effectivealtruism.org/posts/DxfpGi9hwvwLCf5iQ/objections-to-value-alignment-between-effective-altruists#Epistemic_Insularity.
[146] *See e.g.,* Cilliam Crosson, *Apply for Red Team Challenge [May 7 – June 4]*, EFFECTIVE ALTRUISM F. (Mar. 18, 2023), https://forum.effectivealtruism.org/posts/DqBEwHqCdzMDeSBct/apply-for-red-team-challenge-may-7-june-4. Clair Zabel has described the norm within strategic Effective Altruist discussions to include the question, "And

participants would often suggest a round of counterarguments. LPP conducted surveys, held discussions, and wrote a report on the topic of honest feedback and "normalizing being more critical toward each other's work."[147] These norms seemed influential in shaping how participants saw their own work. An LPP member noted that they were hired for a research-focused position "with the goal of advancing longtermism," but explained that "advancing" was meant in a scientific sense, which includes developing counterarguments and identifying uncertainties, and thus "pursuing the truth, be it favorable to longtermism or not…through science and research."

Given this commitment to continually expressing misgivings and objections, it is perhaps surprising that this culture is also marked by what LPP labels a "warm, kind, and supportive" environment, which was very much how this culture appeared through the lens of this qualitative study. For example, weekly all-hands meetings began with a round of "achievements and gratitude" where attendees shared personal updates, e.g., openly disclosing details about health, personal challenges, childcare, pets, family, and favorite TV shows including multiple references to baking shows that were especially appreciated because the contestants are "so kind to each other." In this round of personal updates and at other times, participants openly expressed appreciation for various things in their lives including assistance and feedback they received from their colleagues. LPP members also held regular sessions focused on mental health, team-building activities, "informal hangouts," and one-on-one "watercooler" meetings. When members spoke during online meetings, they were regularly greeted with supportive emojis, e.g. hands clapping, party hat, hearts, and crying laughing.

This norm of kindness and mutual support seemed to play a crucial role in the expression of dissent. Support for a colleague's comment was most often directed toward their reasoning or articulation of different positions, not necessarily their conclusion, which generally remained subject to uncertainty and debate even at the end of a discussion. Some participants suggested that these warm interactions, when paired with a norm of valuing counterarguments, can encourage members to view disagreement as a means to find the best ideas rather than as a personal criticism. Members of the Effective Altruist community sometimes joke about their own linguistic norms, laughing in a self-effacing manner when they say "I claim…I *sub*-claim" as they raise arguments. But this language reveals a serious discursive effort to keep a focus on the content of the claims, which are freely open for debate, rather than focusing on the person making the claims.

It is possible that these efforts to avoid offending one another could lead to a "too nice" dynamic, as one participant noted, such that members of the community could grow unwilling to express dissent out of fear of breaking a norm of amicability. Moreover some commentaries have expressed the opposite concern, suggesting that Effective Altruism, and the existential risk community in particular, are inhospitable to certain lines of counterargument, e.g. regarding the

---

the thing I'm getting wrong about all of this?" Claire Zabel, Program Officer for Global Catastrophic Risks, Open Philanthropy, Fireside Chat at Stanford Existential Risks Conference 2022 (Feb. 27, 2022). *See also* Holden Karnofsky, *Learning by Writing*, LESSWRONG (Feb. 22, 2022), https://www.lesswrong.com/posts/ii4xtogen7AyYmN6B/learning-by-writing.

[147] In some public-facing materials, LPP has described their cultural commitment to "discuss ideas openly and honestly, letting the best ideas win. Honest criticism and feedback are both welcome and expected." LEGAL PRIORITIES PROJECT, *supra* note [X].

value of using the priorities methodology or regarding the common conclusions that existential risk and AI risk in particular should be prioritized.[148] Some have suggested that a monoculture has taken hold in spite of rhetorical efforts to maintain a heterogeneous idea space. But, overall, as observed in this study, this culture of what I reference in my field notes as "supportive dissent" was one of the most striking features of existential advocacy.[149] Over the first few weeks of immersing myself in the existential risk community, I observed repeatedly in my notes that this community's norms of dissent and questioning assumptions and evidence far surpassed what I experience in academia—which is often thought of as a bastion of skeptical inquiry. LPP members and other participants in this study showed a consistent willingness to revisit and debate even the most fundamental concepts that motivate their work, as well as the meta-level question of finding the right mix of agreeableness and dissent.

D. The Epistemic Identity Norm

The participants in this study seemed highly reluctant to identify as part of a larger group of existential advocates. Even full-time employees of organizations in this space were quick to express caveats and reservations when describing themselves as "Effective Altruists," "longtermists," or "advocates for existential risk mitigation."[150] As one participant explained, identifying too strongly with such labels could imply a "movement [with] core tenets you have to follow," whereas they conceived of themselves belonging to a "scientific community" marked by heterogeneous discourse, dissent, and uncertainty. This concept of a scientific community seems designed to limit the sense of group solidarity around a shared identity.[151] Participants suggested that a strong sense of solidarity would be undesirable because it would create "social pressure to conform," a preference for in-group members, and, as one participant put it, an "us vs. them…antipathy" toward out-groups.[152] Similar concerns have been raised by sociologists who

---

[148] See, e.g., Carla Zoe Cremer and Luke Kemp, *Democratising Risk: In Search of a Methodology to Study Existential Risk* (forthcoming). *See also* Lucius Caviola, *Against Naïve Effective Altruism* (Nov. 20, 2017), https://www.youtube.com/watch?v=-2oRgxxafXk&ab_channel=CentreforEffectiveAltruism (describing the risk that a naïve understanding of Effective Altruism could lead to being overly concerned with "signaling contrarianism" and "appearing cool and wise by holding weird beliefs," and as a result being overly dismissive of common sense).

[149] While writing this article, I learned that this term is very similar to what Effective Altruists call, "supportive scepticism in practice." *See* Michelle Hutchinson, *Supportive Scepticism in Practice*, https://forum.effectivealtruism.org/posts/CkikpvdkkLLJHhLXL/supportive-scepticism-in-practice. Michelle_Hutchinson, *Supportive Scepticism in Practice*, EFFECTIVE ALTRUISM F. (Jan. 15, 2015), https://forum.effectivealtruism.org/posts/CkikpvdkkLLJHhLXL/supportive-scepticism-in-practice.; Lizka, *Guide to Norms on the Forum*, *supra*, https://forum.effectivealtruism.org/posts/yND9aGJgobm5dEXqF/guide-to-norms-on-the-forum (advising that "when you criticize someone's point, consider doing so supportively."). *See also t*he EA Forum Guidelines ("when you criticize someone's point, consider doing so supportively.").]

[150] Note that "existential advocacy" is a novel term of analysis used in this article. This is not a term circulating among advocates in this field.

[151] One participant noted that Effective Altruism does not particularly lend to identification because it is a "very diverse movement" full of "different views and priorities" and a strong commitment to debate. *See also*, Ajeya Cotra, *supra* note ("I say that I'm an Effective Altruist. That just means a person trying to be effective at altruism.").

[152] See Benjamin Todd (@ben_j_todd), Twitter (Aug. 8, 2021, 3:54PM) https://twitter.com/ben_j_todd/status/1424458937286512647 ("turning Effective Altruism into an identity has been powerful, but has had many downsides," and this includes that it "creates social pressure to conform"); Lizka, *Against Longtermist as an Identity*, EA FORUM (May 13, 2022) https://forum.effectivealtruism.org/posts/FkFTXKeFxwcGiBTwk/against-longtermist-as-an-identity (arguing that identifying as a longtermist can "make it harder to change your mind based on new information" and can lead to

find that members of social movements tend to develop a strong preference for in-groups, while undergoing "identity work" to "preserve or to enhance their egos" by aligning their views with the goals of the movement.[153] While such a solidarity process may make movements more effective in some respects, it may tend to limit available "argument pools" due to dynamics of "enclave deliberation."[154] Moreover, participants in this study commonly cited psychological research on "moral tribes" and "righteous minds,"[155] widely read in the Effective Altruism community, suggesting that when groups band together around certain viewpoints they tend to falter in their truth-seeking and moral reasoning. Yet, a recurring theme in my field notes was the surprising (to me) absence of judgment and derision toward out-groups. I had anticipated that this community that is working to set optimal priorities might view the rest of the world as setting the "wrong" priorities (e.g. failing to address existential risk) and thus would tend to define themselves in opposition to these other groups. But over the course of this study such judgment was rarely expressed and did not appear to be a strong source of motivation.

As represented in this study, the members of this community seek to limit but not entirely reject group identification. Participants acknowledged some benefits of aligning personal and collective identities—similar to the "identity/movement nexus" cited by sociologists as a crucial element in the formation and growth of social movements.[156] Some participants described their

---

confusion where the "group identity" encourages viewpoints that one would "otherwise not have adopted as part of your individual identity."); Helen *Effective Altruism is a Question (not an ideology)*, EA FORUM (OCT. 16, 2014) https://forum.effectivealtruism.org/posts/FpjQMYQmS3rWewZ83/effective-altruism-is-a-question-not-an-ideology (arguing against identifying as an Effective Altruist in favor of stating that one is an "aspiring Effective Altruist," a "member of the Effective Altruism movement," or "interested in Effective Altruism"); *see also* Cullen O'Keefe, *Proposed Longtermist Flag*, EA FORUM, https://forum.effectivealtruism.org/posts/efd4B2LLd3DXGivSP/proposed-longtermist-flag (Mar. 24, 2021) (suggesting the possibility of creating a flag for longtermism, as has proven useful in "mature and successful movements," but raising the concern that a flag might encourage partisan psychology and "ideological loyalty," and that it could "stymie open and honest discourse about longtermism, including criticism thereof.").

[153] CANTRIL, *supra* note at 162 (observting that, upon joining a social movement, "the individual is now an in-group member of a rather highly selected gathering," and that such identification with the movement can cause the individual to "lose themselves in some causes"); Jasper, *supra* note (suggesting that movements tend to "minimize affective loyalties to anyone outside the group and maximize them to the group or its leaders," and that movements tend to frame in-groups positively relative to out-groups).

[154] SUNSTEIN, *supra* note (observing that when one conceives of oneself "as part of a group having a degree of connection and solidarity…group polarization is all the more likely, and it is also likely to be more extreme…" and noting that these dynamics "tend to suppress dissent and thus to lead to inferior decisions.").

[155] *See generally,* Joshua Greene, MORAL TRIBES: EMOTION, REASON, AND THE GAP BETWEEN US AND THEM (2013) (discussing the human tendency to rely on "tribal gut reactions" and our "automatic" rather than "manual mode," which raises a "distinctively modern moral problem"); JONATHAN HAIDT, THE RIGHTEOUS MIND: WHY GOOD PEOPLE ARE DIVIDED BY POLITICS AND RELIGION (2012) (detailing the psychological bases for the human tendency toward group-based morality, which leads to a failure to appreciate valid perspectives among out-groups).

[156] Snow and Mcadam, *supra* (citing the example of the mid-twentieth century Civil Rights Movement, where movements leaders drew on church identity, in addition to racial identity, to mobilize and foster commitment to the cause, while also transforming identities of activists in the Freedom Summer). *See* DAVID SNOW, COLLECTIVE IDENTITY AND EXPRESSIVE FORMS (2001) (discussing mechanisms of identity convergence in social movements, including amplification (building on existing identities that are already congruent with the movement), consolidation (revealing the compatibility of identities that seemed inconsistent), extension (making existing identities more pervasive or salient), and transformation (forming new identities)); CASS SUNSTEIN, HOW CHANGE HAPPENS 2019 (noting that group polarization has been useful to spur movements, citing a number of examples including feminism, the Civil Rights Movement, and Reaganism); Eric Neyman, *Can Group Identity be a Force for Good?*,

affinity with fellow existential advocates in terms of a common "life plan,"[157] a set of background readings, and a general commitment to "thinking about making a fair world" through an evidence-based methodology.[158] This creates a basis for some degree of shared identity, although this identity is rooted primarily in their commitment to their methodology and epistemic culture rather than any particular conclusions about what cause areas or strategies should be most prioritized. As one participant reflect: "We unite around our epistemic norms." These accounts of what I referred to in my field notes as "epistemic identification" were near ubiquitous in my interviews and ethnographic observations.

This approach to group identity may not be the most powerful basis for motivation and recruitment. One participant described their relatively "weak" identification as an Effective Altruist and contrasted this with their past experience in movements where they attended protest demonstrations for criminal justice reform and other "leftist" causes, which fostered feelings of "certainty and a sense of belonging and having a clear enemy." Without a clear enemy or even a clear out-group, it is perhaps more difficult to decipher what one's collective identity is—to demarcate where one's identity begins and ends. Moreover, epistemic identification may fail to provide the same "warm glow," as this participant put it, of belonging and collective voice. But this same participant noted that while they were engaged in other social movements, what was missing was a "focus on reasoning," which they were grateful to find in the existential advocacy community. Other participants similarly observed that they had generally been disappointed by the lack of evidence-based reasoning and high epistemic standards in their experiences in academia, legal practice, government service, and in society generally. As already noted, these participants did not tend to denigrate these other contexts. Instead, they emphasized their sense of appreciation for the existential risk community (and the related Effective Altruist and longtermist communities). Some participants offered biographical accounts of first discovering these communities and "finding a home" when they realized the shared commitment to, as one participant put it, "take epistemics seriously."

In sum, this analysis suggests a tension within the existential advocate identity—which offers a sense of belonging but also demands that one resist overly identifying with the group. Several participants noted that the shared experience of this tension can foster a sense of solidarity, whereby members identify with one another on the basis of the shared norm against over-identification. This form of solidarity is grounded in a contradiction (group identity based on resisting group identity), and so is inherently limited. But perhaps this is to a desired degree. Some participants suggested that maintaining this identity tension is crucial to the cultural underpinnings of the priorities methodology. Too little identification with movement and cause

---

UNEXPECTED VALUES (July 4, 2021), https://ericneyman.wordpress.com/2021/07/04/can-tribalism-be-a-force-for-good.

[157] *See* Ajeya Cotra, *supra* note (noting that Effective Altruism presents a "radical but very simple plan for your life. Figure out how to do the most good and then do it."). LEGAL PRIORITIES PROJECT, *supra* note [X] (noting that the LPP project is the "main life project" for most members of the team).

[158] This common set of readings and other Effective Altruism materials was evident, for example, in the application questions for a recent reading group on law and Effective Altruism, which asked if applicants had attended Effective Altruist events, listened to relevant podcasts, read career advice from 80,000 Hours, read the Effective Altruism Handbook, or read writings from a list of Effective Altruist and longtermist theorists, including Will MacAskill, Nick Bostrom, and Toby Ord.

might lead to a lack of motivation. But too much identification might tend to undermine the cultural commitments to uncertainty, dissent, and deliberative decision-making.

VI. Discussion

In his 2004 book, *Catastrophe: Risk and Response*, Professor Richard Posner argued that the legal profession had failed to play its crucial role in mitigating catastrophic risks, including risks on the existential scale, and that this was largely due to the profession's lack of scientific methodology and mathematical reasoning.[159] This criticism took particular aim at lawyers, who, owing to their "culture of advocacy and doctrinal manipulation," tend to assume a truth favorable to their client and then seek to persuade others of that truth rather than engaging in a scientific process of evaluating which claims are most true.[160] Posner contrasts the scientific orientation "toward knowledge," as demonstrated through experimental research, with the law's orientation "toward action," "bending the rules…fitting them to goals," and asserting certitude, by, for example, declaring "my client is innocent, and that's the truth."[161] Posner advised law schools to recruit more students with STEM backgrounds and provide basic STEM education within the law curriculum. This would then help produce a class of "catastrophic risk lawyers" who appreciate the prerequisite scientific methods, probabilistic claims, and decision theory, that is, lawyers who recognize that uncertain, low-probability harms can be deserving of our legal attention when the harms carry a great magnitude.

The participants in this study are living evidence of the exceptional case that Posner hoped law schools would foster—a group of lawyers who are addressing global catastrophic risks through an express commitment to scientific reasoning and truth-seeking. These advocates generally lack the STEM backgrounds emphasized by Posner. Instead, they find a commitment to scientific reasoning in the priorities methodology drawn from the framework of Effective Altruism. Posner's analysis would seem to suggest a culture clash between the notion of learning to "think like a lawyer"[162] and what MacAskill calls learning to "think like an Effective Altruist."[163]

---

[159] POSNER, *supra* note (observing that law lacks "empirical methodology" and "error correction," and is "more like a language than a science").

[160] POSNER, *supra* note. This concern is reflected in the laments of legal empiricists that the profession disregards empirical research. *See* James D. Greiner, *The New Legal Empiricism & Its Application to Access-to-Justice Inquiries,* 148.1 DAEDALUS 64 (2019) (arguing that the legal profession is "not evidence-based in the scientific sense" and instead tends to "rely on gut intuition and instinct, not on rigorous evidence," and advocating for the "new legal empiricism," which has the potential to "transform the U.S. legal profession into an evidence-based field"). This concern also finds support among critics of the role-differentiated morality of lawyers' professional identities, as the professional role demands zealous partisan advocacy for clients with relatively little opportunity for the lawyer to insert their own political, moral, emotional, and other contextual considerations. This "hired gun" loyalty to clients is limited by professional duties to the public and a blanket requirement to not engage in dishonesty or perjury, but these duties are rarely interpreted to suggest anything like a commitment to scientific methods or reasoning. *See generally* Richard Wasserstrom, *Lawyers as Professionals: Some Moral Issues*, 5.1 HUMAN RIGHTS 1, 23 (1975);

[161] POSNER, *supra* note ("the idea of subjecting a legal proposition to a decisive experiment…horrifies the lawyer.").

[162] ELIZABETH MERTZ, THE LANGUAGE OF LAW SCHOOL: LEARNING TO "THINK LIKE A LAWYER" 226 (2007) (detailing the amoral, apolitical, acontextual, and unemotional dimensions of legal epistemology as communicated through law school teaching).

[163] MACASKILL, *supra* note.

Similarly, scholars of professional responsibility may find the priorities methodology a surprising fit for a community of lawyers. Although lawyers are obligated to protect the legal system and the general public, the core mandate of their professional role is to serve as a partisan advocate for clients. The tradition of cause lawyering, as discussed supra (Part IV), places a greater focus on advancing legal and social change. But even within this tradition, the notion that lawyers would find their deepest sense of accountability in a formal methodology for maximizing impact is anomalous. How have these lawyers become what we might label, "maximizing lawyers" or "prioritist lawyers," zealously committed to scientific reasoning and net moral impact?

One answer may be that the scientific model used by these advocates is a function of privileged identities. This community is disproportionately white, male, and elite educated. It is remarkably diverse in terms of representing different nationalities, but there is a tilt toward the Global North.[164] The prevalence of these privileged identities may tend to give this community less appreciation for current suffering and oppression in the world, leading to a focus on other populations, namely those who will exist in the future.[165] These privileged identities could also help to explain some of the cultural traits outlined in this article—efforts to enhance rational deliberation while limiting group identity and emotional reasoning.

These speculations about how privileged identities influence the practice of existential advocacy should be subjected to empirical inquiry, but if these identify effects are significant then it is all the more important that this movement continues its efforts to diversify its ranks. As discussed supra part IV, diversification is valued by advocates in this field because it may lead to greater inclusivity and legitimacy, as well as more accurate assessments of prioritized causes and strategies. Recent scholarship on public-interest law has noted that the prevalence of white leadership in movements for civil rights has led these movements to overlook key injustices and other crucial considerations.[166] Similarly, the existential risk community would likely benefit from more perspectives of people who experience different forms of injustice and who live relatively dystopic lives today—analogous to the permanent dystopian scenarios contemplated by scholars of existential risk.

The effort to create a scientific methodology to maximize de facto impact may also be influenced by the nature of existential risk as a cause area, which is an unusual topic to serve as the subject of a movement for legal and social change. The empirical literature on social-change lawyering has primarily focused on grassroots social movements that address matters of public controversy and pervasive cultural norms (e.g., how we relate to one another across lines of race, gender, and sexuality). In contrast, the advocates examined in this article are addressing an issue that is not particularly affected by the general population's daily norms and culture. Moreover, even if these advocates were to attempt to form a broader social movement, it may be exceedingly difficult to

---

[164] See Supra introduction.

[165] It is important to note here that participants in this study seemed remarkably attentive to issues of present-day global health and poverty, which continue to be the cause areas where the most philanthropic dollars are spent in the Effective Altruism community.

[166] *See* Atinuke O. Adediran & Shaun Ossei-Owusu, *The Racial Reckoning of Public Interest Law*, 21 CAL. L. REV. (2021) (calling for greater scrutiny of the racial composition of U.S. public interest law as it impacts marginalized communities).

mobilize large populations around low-probability/high-impact harms, particularly where such harms are framed as a threat primarily to a distant population of future generations. This lack of what socio-legal scholars call "mobilizing frames" may help to explain why these advocates are attracted to the priorities methodology.[167] If this community remains a relatively small (but hopefully diversifying) group of advocates dedicated to working full-time on the complex strategic considerations around existential risk, this may be the sort of community that is well-positioned to take a scientific approach to setting priorities and designing strategies.

While this notion of maintaining a small expert-based movement holds a great appeal for many of these advocates, this community has recently stepped into a very high-profile public spotlight. Existential risk had already been the subject of public-facing books, videos, blogs, and podcasts,[168] as well in-depth New Yorker profiles of the leading scholars.[169] But 2022 saw the field truly, and quite suddenly, enter the mainstream of popular and political debate. This was the product of three factors in particular. First, in August 2022, the release of William MacAskill's book, *What We Owe the Future*, was a New York Times Bestseller and received highly favorable coverage across nearly all leading news outlets, including a Time Magazine cover story.[170] Second, just two months later, the leading funder of existential advocacy, Sam Bankman-Fried, who had pledged upward of $30 billion to the field, became the biggest corporate fraud news story in years. Much of the public outcry about Bankman-Fried took aim at his philanthropic and political efforts, with many journalists and academics ridiculing existential risk as a tech billionaire fantasy—or, even worse, as a misconceived public relations effort to bolster the

---

[167] *See* David A. Snow, Rens Vliegenthart, and Pauline Ketelaars, *The Framing Perspective on Social Movements: Its Conceptual Roots and Architecture*, 77 THE WILEY BLACKWELL COMPANION TO SOCIAL MOVEMENTS 392 (2019).

[168] *See generally*, ORD, *supra* note; BOSTROM, *supra* note; MACASKILL, supra; Karnofsky, *supra*; Nick Bostrom, *The End of Humanity*, YOUTUBE (MAR. 26, 2013), https://www.youtube.com/watch?v=P0Nf3TcMiHo&ab_channel=TEDxTalks; *The Last Human: A Glimpse into the Far Future*, YOUTUBE (June 28, 2022) https://www.youtube.com/watch?v=LEENEFaVUzU&t=318s&ab_channel=Kurzgesagt%E2%80%93InaNutshell; FUTURE OF LIFE INST., https://futureoflife.org/the-future-of-life-podcast (last visited Aug. 7, 2022); *See also* VOX: FUTURE PERFECT, https://www.vox.com/future-perfect (last visited Aug. 7, 2022).

[169] *See* Corinne Purtill, *How Close is Humanity to the Edge?*, NEW YORKER (Nov. 21, 2020), https://www.newyorker.com/culture/annals-of-inquiry/how-close-is-humanity-to-the-edge; Raffi Khatchadourian, *The Doomsday Invention*, NEW YORKER (Nov. 23, 2015), https://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom); Gideon Lewis-Kraus, *The Reluctant Prophet of Effective Altruism* NEW YORKER (Aug. 15, 2022) https://www.newyorker.com/magazine/2022/08/15/the-reluctant-prophet-of-effective-altruism.

[170] *See, e.g.,* William MacAskill, *The Case for Longtermism*, THE NEW YORK TIMES (Aug 5, 2022) https://www.nytimes.com/2022/08/05/opinion/the-case-for-longtermism.html; *Three Sentences that Could Change Your Life,* THE EZRA KLEIN SHOW (interview with Will MacAskill) (https://www.nytimes.com/2022/08/09/opinion/ezra-klein-podcast-will-macaskill.html) (Aug. 9, 2022); Naina Bajekal, *supra* note 17; *William MacAskill – "What We Owe The Future*, COMEDY CENT.: THE DAILY SHOW WITH TREVOR NOAH (Sept. 9, 2022), https://www.cc.com/video/8fl6g9/the-daily-show-with-trevor-noah-william-macaskill-what-we-owe-the-future; but see criticism of *What We Owe the Future* in *But see* Christine Emba, *Opinion: Why 'Longtermism' Isn't Ethically Sound*, WASH. POST (Sept. 5, 2022), https://www.washingtonpost.com/opinions/2022/09/05/longtermism-philanthropy-altruism-risks; Alexander Zaitchik, *The Heavy Price of Longtermism*, NEW REPUBLIC (Oct. 24, 2022), https://newrepublic.com/article/168047/longtermism-future-humanity-william-macaskill; Kieran Setiya, *The New Moral Mathematics*, BOS. REV. (Aug. 15, 2022), https://www.bostonreview.net/articles/the-new-moral-mathematics.

business interests of the greatest fraudster of a generation.[171] Third, the November 2022 release of ChatGPT brought public awareness to transformative advances in artificial intelligence, leading to a new appreciation for the strange and potentially dangerous world of unprecedented technology we are entering.[172] These developments build on other events that have brought public attention to existential risk in the past few years, such as COVID-19 generating near ubiquitous discussion of global pandemics, the invasion of Ukraine raising fears of autonomous weapons and nuclear war, and extreme fires, floods, and temperatures bringing a new urgency to the worst climate change scenarios.[173]

A non-profit research organization was recently founded to investigate whether the existential risk community should pursue broader movement-building strategies.[174] Drawing on empirical research and analogous case studies, the organization strongly recommended the formation of social movement organizations, which would organize protest demonstrations and other actions intended to shape public opinion and put pressure on key decision-makers. It is not clear yet whether advocates in this field will pursue these grassroots strategies—and whether they would succeed in developing a larger movement if they tried. What is clear is that the cat is out of the bag and a significant degree of public engagement around the issue of existential risk is now unavoidable.

The scholarship on law and social change would seem to generally support the notion of expanding the community of existential advocates. As discussed supra Part I.B, scholars in this field overwhelmingly recommend "integrated advocacy," noting that legal activists in other social movements have found greater efficacy and accountability by embedding their legal work within larger movements and seeking to shape public opinion.[175] This approach is thought to help promote lasting social change by creating a collective demand for reform, which motivates favorable legal and political decision-making and diminishes backlash to legal victories. Just how far should the existential risk community take this received wisdom from the academic literature? Should they recruit new members with the widest possible net, expanding their movement to create a larger collective voice? Should they focus on public opinion, seeking to foster a world where people generally understand the notion of existential risk and would support interventions as opposed to viewing the issue as a matter of science fiction or billionaires' whims? Or should this community stay relatively narrow, working to put experts in conversation with powerful decisionmakers (e.g. lawmakers and judges)?

---

[171] *See* supra note 32.

[172] Kelsey Piper, *ChatGPT has Given Everyone a Glimpse at AI's Astounding Progress*, VOX (Dec. 15, 2022), https://www.vox.com/future-perfect/2022/12/15/23509014/chatgpt-artificial-intelligence-openai-language-models-ai-risk-google (noting that ChatGPT, a chatbot drawing from a large language model, is "the general public's first hands-on introduction to how powerful modern AI has gotten[.]");

[173] Amy Maxmen, *Has COVID Taught us Anything About Pandemic Preparedness?*, NATURE (Aug. 13, 2021), https://www.nature.com/articles/d41586-021-02217-y (noting increased awareness of global pandemic risks in the wake of COVID-19 but a general lack of response among lawmakers); *see also* Bridget Williams & William MacAskill, *Investing in Pandemic Prevent is Essential to Defend Against Future Outbreaks*, BULL. OF THE ATOMIC SCIENTISTS (Nov. 2, 2022), https://thebulletin.org/2022/11/investing-in-pandemic-prevention-is-essential-to-defend-against-future-outbreaks.

[174] https://www.socialchangelab.org/

[175] *See,* CUMMINGS, *supra* note.

These are matters for further research beyond the scope of this article, but one consideration bears directly on this article's core findings: Efforts to expand this movement may tend to compromise the culture underlying the priorities methodology. The "mobilizing frames" that social scientists have identified as the key ingredients for building a broad social movement are seemingly point-by-point the exact opposite of the cultural commitments outlined in this article. For example, recruiting more broadly may require a more conventional approach to identifying with movement, cause, and strategy. Most social movements depend on some group taking identity ownership over an issue (e.g. young people could include existential risk within the climate youth movement). In contrast, the existential advocates, under their priorities model, seek to limit these identity dynamics. Moreover, efforts to persuade broader populations may require presenting existential risk mitigation in more familiar terms, meeting people where they are, putting less emphasis on uncertainty, and making issues more emotionally salient. The affective dimension could be enhanced by stirring up fear and other strong emotions around existential risk. Some organizations in this field have explored this approach by producing short films about "slaughterbot" scenarios, where swarms of small autonomous drones execute people in great numbers. Participants in this study tended to disfavor this approach on the grounds that, as discussed above (supra Part V.3), motivations rooted in powerful emotions may cause the movement to drift toward a focus on near-term and smaller-scale catastrophic risks, while losing sight of existential threats to the future of humanity.

At least at this early stage, the existential advocates seem very effective at avoiding this drift. They are remarkably consistent in their methodology and their cultural commitments, assessing nearly all strategic decisions according to how much each option is expected to reduce our overall level of existential risk. This ability to "keep your eyes on the prize," in the refrain that echoed through the Civil Rights Movement,[176] is a defining feature of the social-change lawyering observed in this study. As this community scales up and pursues more direct legal interventions, their culture will likely need to adapt and compromise. But maintaining some degree of relatively uncompromised commitment to the priorities methodology could be vital to this movement's success. As humans, we carry cognitive biases that limit our ability to appreciate existential risk, an issue that is uncertain, unprecedented, large-scale, and seemingly remote—affecting future generations who are an abstraction well beyond our usual moral circles. Moreover, legal and political systems have incentives to focus on issues of more immediate concern to voters, markets, and movements. Creating a legal movement dedicated to overcoming, or at least challenging, these biases and incentives is no small task. These advocates are drawing on a novel framework of continual evidence-based analysis backed by a culture of scientific reasoning. In this way, they aim to keep their focus on representing the current and future generations whose well-being and existence hang in the balance.

[176] This is a lyric from the traditional African American spiritual, "Gospel Plow." *See* DUKE ELLINGTON & MAHILIA JACKSON, *Keep Your Hand on the Plow*, *on* LIVE AT NEWPORT 1958 (Columbia Records 1958).